Expression and characterization of integrase from Moloney murine leukemia virus

Koepke, Kristine Ann

Award date:
1994

Awarding institution:
University of Bath

EXPRESSION AND CHARACTERIZATION OF INTEGRASE FROM MOLONEY MURINE LEUKEMIA VIRUS

submitted by Kristine Ann Koepke

for the degree of PhD

University of Bath 1994

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Integrase (IN) is the retroviral enzyme that is responsible for integrating the viral DNA into the hosts genome. IN processes the viral DNA by cleaving two nucleotides from the 3' ends, then cleaving the host DNA and joining the viral and host DNA. IN from Moloney murine leukaemia virus (MoMLV) was expressed in *E. coli* using two different expression systems and characterized in several *in vitro* assays.

The first expression system produced IN in insoluble inclusion bodies. Denaturation and refolding trials were carried out; soluble IN was produced by solubilization in 6 M guanidine and dialysis against decreasing concentrations of urea. To aid in the identification of IN during these trials, two synthetic peptides were synthesized and used to produce immune sera.

The second expression system, as a fusion with maltose binding protein (MBP), produced soluble fusion protein, but IN precipitated out of solution when cleaved from MBP. The hypothesis was put forward that the improper folding of IN is due to high proline content.

IN was tested in a variety of *in vitro* assays, including a novel, continuous, direct kinetic assay. In each case, IN was found to be inactive. In was also tested for zinc affinity and found to bind zinc. Sensitivity to oxidation and competition studies suggested that zinc is binding to a zinc finger motif at the N-terminus of the protein.

Acknowledgments

Many thanks to Adrian, Mike, and Dave for all their support, guidance, and enthusiasm during the project. To Drs Janet Rider, Tony Else, and Alan McEuen for all their help and advice. To all the people in the End Lab, Jackie, Michelle, Rupert, Keith, Bryan, that made life so enjoyable and bench work fun.

Thanks to Alan McEuen for developing the zinc blotting assay and doing the first zinc blots with MoMLV IN. Many thanks to Richard Kinsman for the peptide synthesis and the humorous times in his lab. Thanks to Karen Roberts who did the trial purifications of MBP-IN and the SPA assays on MBP-IN. Thanks to Amersham, Cardiff for the SPA reagents and oligos and special thanks to Malcolm Downes for advice and helpful discussions.

Special thanks to The Wellcome Trust for the Wellcome Prize Studentship that funded this project.

And finally, many thanks to Lucas Partridge, for support, motivation, helpful discussions, and for taking me away from the world of molecules to introduce me to the joy of ants on Portland Bill.

Abbreviations

AIDS Acquired Immune Deficiency Syndrome
ALV Avian Leukosis Virus
AMV Avian Myeloblastosis Virus
BCIP 5-bromo-4-chloro-3-indolyl phosphate
BLV Bovine Leukaemia Virus
CA Capsid protein
CDM Counterion Distribution Monitoring
DEPC Diethyl pyrocarbonate
DMF Dimethyl formamide
DTPA diethylenetriaminepentaacetic acid
DTT Dithiothreitol
EDTA Diaminoethanetetra-acetic acid
EGTA Ethyleneglycobis(aminoethyl)-tetra-acetic acid
ELISA Enzyme Linked Immunosorbent Assay
HEPES 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HIV Human Immunodeficiency Virus
HTLV Human T-cell Leukaemia Virus
IN Integrase
IPTG Isopropyl-β-D-thiogalactopyranoside
LTR Long Terminal Repeat
MA Matrix protein
MBP Maltose Binding Protein
MBP-IN Maltose Binding Protein-Integrase fusion protein
MBS Maleimidobenzoic acid N-Hydrosysuccinimide ester
MLV Murine Leukaemia Virus
MoMLV Moloney Murine Leukaemia Virus
NBT Nitro Blue Tetrazolium
NC Nucleoprotein
PAGE Polyacrylamide Gel Electrophoresis
pbs primer binding site
PIPES Piperazine-1,4-bis(2-ethanesulfonic acid)
PNK Polynucleotide kinase
PPI Peptidyl-prolyl-*cis-trans*-isomerase
RP-HPLC Reverse Phase-High Performance Liquid Chromatography
RSV Rous Sarcoma Virus

RT Reverse Transcriptase
SDS Sodium dodecyl (lauryl) sulfate
SIV Simian Immunodeficiency Virus
SPPS Solid Phase Peptide Synthesis
SU Receptor binding protein
β-ME β-mercaptoethanol
TCA Trichloroacetic acid
TEMED N,N,N',N'-tetramethyl-ethylenediamine
TFA Triflouroacetic acid
TM Transmembrane protein
Tris Tris(hydroxymethyl)aminomethane

Table of Contents

Chapter 1

1.0 Introduction

1.1 Retroviruses

The *Retroviridae* constitute a family of viruses with two unique events in their replication cycle. The first is the synthesis of a dsDNA copy of the ssRNA genome; the second is the integration of that DNA into the host genome.

Retroviruses infect a wide variety of vertebrates. Examples include the murine leukaemia viruses, avian sarcoma/leukosis viruses, feline leukaemia viruses, human T-cell lymphoma viruses and the human immunodeficiency viruses types 1 and 2 (reviewed by Varmus and Brown, 1989; Coffin, 1990).

1.2 Classification and Pathology

Retroviruses are divided into three genera: oncovirinae, lentivirinae and spumavirinae. Retroviruses can be described by virion structure and assembly. Type A particles are found only intracellularly, and are double shelled. Type B are larger than Type A and doughnut shaped with surface spikes. Type C are associated with budding at the cell's plasma membrane. Type D are assembled intracellularly and delivered to the plasma membrane (Rhee and Hunter, 1990) having shorter spikes than Type B. Another method of

description is by overall genome organization and species of origin giving seven major groups: avian leukosis-sarcoma, mammalian type-C, type-B, type-D, HTLV-BLV, lentiviruses and "foamy" viruses.

Retroviruses display a variety of pathological effects: nonpathogenic, weakly oncogenic, highly oncogenic or cytopathic (reviewed by Temin, 1988). The cytopathological effects of proviral insertion depend upon the integration site. The provirus can physically interrupt and inactivate cellular genes, or the regulatory features of viral DNA may alter the expression of neighbouring genes, or the provirus can recombine with a cellular gene that is implicated in viral oncogenesis. Insertion of a provirus near cellular oncogenes has been demonstrated by avian, murine and feline retroviruses. Most retroviruses do not kill infected cells, rather a growth advantage is conferred upon the cell.

1.3 Murine Leukaemia Viruses (MLV) and Moloney Murine Leukaemia Virus (MoMLV)

The murine leukaemia viruses are divided into subgroups by their host range determinants. Ecotropic viruses infect only murine cells and include Moloney murine leukaemia virus and Friend murine leukaemia virus. Amphotropic viruses also infect murine cells but their range is extended and they can infect cells from other species. Xenotropic viruses are found in the germline of inbred mice but do not infect mouse cells from these strains. A fourth class is the 10A1 virus,

which is a variant amphotropic virus with enhanced leukemogenicity and expanded host range from its amphotropic parent.

The pathogenicity of MLVs are varied. AKV is non-leukemogenic, SL3-3, MoMLV and MCF's cause T-cell lymphomas, while Friend MLV is erythroleukemogenic.

MoMLV was first described by Moloney (1960) after recovery of the virus from a sarcoma. Since then, the genome of MoMLV has been cloned and sequenced (Shinnick, 1981). MoMLV belongs to the Oncovirinae subfamily and is an exogenous Mammalian Type-C virus.

Pathogenesis by MoMLV is a two step process with a latency of several months. The first stage is a mild splenomegaly and general haemetopoeitic hyperplasia. The second stage is a secondary infection of hyperplastic lymphoid stem cells followed by migration to the thymus. It is the second event that results in site specific insertion and activation of a cellular proto-oncogene (Cuypers *et al.*, 1984; Davis *et al.*, 1987; Storch *et al.*, 1985). Leukemogenesis is associated with proviral amplification and selection for cells containing multiple copies of the proviral DNA (Tsichlis, 1987). The chromosomal integration sites have been mapped by several groups with common sites occurring. Tsichlis (1987) demonstrated 4 sites, *Mlvi 1*, *2* and *3* (Murine leukaemia integration) as well as *c-Myc* (confirmed by Steffan, 1984). Integration at Pim-1 (Proviral integration site MLV), found by Cuypers *et al.* (1984) is accompanied by transcriptional activation of

neighboring host sequences. Recent evidence by Makris *et al.* (1993) indicates that MoMLV may pick up a cellular oncogene, recombine, and then integrate.

1.4 The Virus Particle

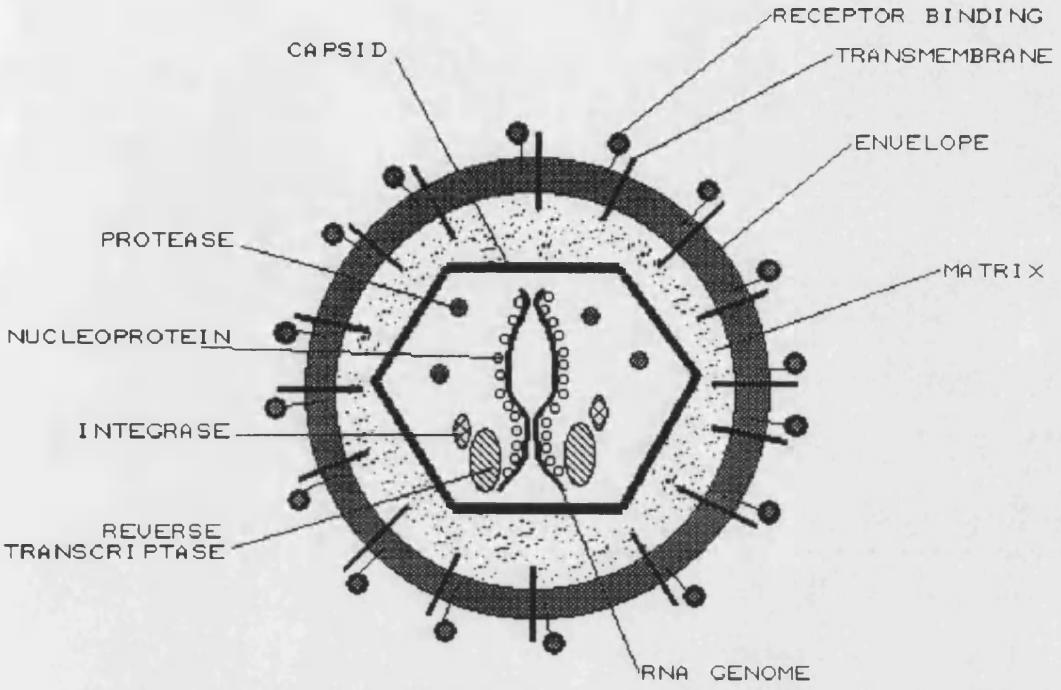
1.4.1 The Genome

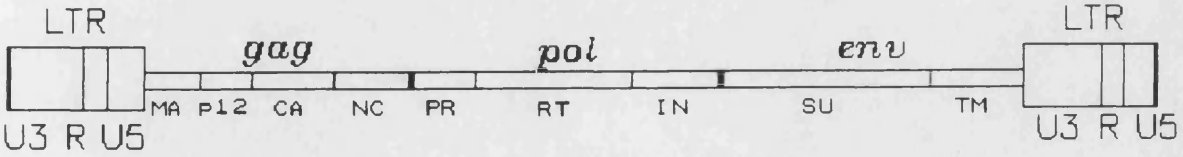
The genome of MoMLV consists of an 8.3 kb plus-stranded RNA with the usual retroviral sequence of coding regions: *gag*, *pol*, and *env* (figure 1.1b). Some eukaryotic features are present: the 5' end is capped, there is a poly-(A) tail at the 3' end and some A residues are methylated. The genome is diploid in that there are two copies, usually identical, associated by hydrogen bonding in a parallel arrangement. The reason for duplication is unknown, but it may allow for recombination and repair (Coffin, 1990).

During reverse transcription the long terminal repeats (LTRs) are created. The LTRs have a great effect on the pathogenicity of the virus [reviewed by Fan (1990) and Majors (1990)] as well as having all the control sequences for the initiation of transcription and polyadenylation. The LTR consists of three regions; U3, R, and U5. U3 (unique 3') contains the promoter for RNA polymerase II with TATA and CAT boxes as well as a 75 bp enhancer. The enhancer is implicated in determining viral pathogenicity, disease specificity and latency period (Chatis, 1983; Golemis,

Figure 1.1a The Retrovirus Particle. A highly schematic diagram of a typical retrovirus particle. The viral proteins are described in detail in sections 1.4.2 and 1.4.3.

Figure 1.1b The MoMLV Genome. Schematic representation of the organization of the viral genes. The genome is described in detail in section 1.4.1.


RECEPTOR BINDING
CAPSID
TRANSMEMBRANE
ENVELOPE
PROTEASE
MATRIX
NUCLEOPROTEIN
INTEGRASE
REVERSE
TRANSCRIPTASE
RNA GENOME


LTR
gag
pol
env
LTR
MA P12 CA NC PR RT IN SU TM
U3 R U5
U3 R U5

1990). Six distinct nuclear factors have been identified which bind to the enhancer (Speck and Baltimore, 1987). The R (repeat) region is involved in mediating minus-strand strong-stop DNA transfer during reverse transcription and has been implicated in controlling LTR-driven expression. The U5 (unique 5') region is essential for initiation of reverse transcription and forming the *att* site which is important in integration. Located 3' to the LTR are the primer binding site (pbs) for the initiation of reverse transcription, the donor splice site for the formation of *env* mRNA, and the *psi* domain, a dimerization and packaging signal that forms a highly structured arrangement of stem loops.

The *gag*, *pol* and *env* reading frames code for all the viral proteins. The placement of the protease reading frame is variable in retroviruses, found in *gag* in the avian type-C viruses and in *pol* in MoMLV (Yoshinaka *et al.*, 1985).

1.4.2 Capsid

The major capsid proteins are encoded by the *gag* region. The *gag* reading frame is translated as a polyprotein precursor, Pr65, and is cleaved by the viral protease during assembly to form the mature proteins: MA (p15), p12, CA (p30), and NC (p10). NC is a basic phosphorylated protein found in association with the genomic RNA. NC has well-conserved Cys-His boxes resembling classical zinc fingers (discussed in greater detail in Chapter 7) and may be

involved in specific recognition of RNA by positioning the tRNA on the pbs (Prats *et al.*, 1988; Housset *et al.*, 1990) and the non-specific function of coating the genomic RNA. CA is hydrophobic and is the major structural and antigenic component. CA has a putative nuclear targeting signal (Nash, 1993) and functions in virion assembly. p12's function is not yet determined. MA is in close association with the envelope, is myristylated at the N-terminus and functions to stabilize the gag polyproteins during assembly (Rhee and Hunter, 1990). The myristate is required for membrane association of $Pr65^{gag}$ during assembly (Rein *et al.*, 1986).

The viral enzymes are also located within the capsid: the protease (reviewed by Skalka, 1989) functioning in cleavage of the polyprotein precursors, reverse transcriptase, an RNA-dependent DNA polymerase that also exhibits RNase H activity, and integrase, mediating the insertion of the viral DNA into the host DNA. The viral particle also contains host derived elements from random packaging. Among these is the $tRNA^{Pro}$ that primes reverse transcription.

1.4.3 Envelope

The envelope itself is derived from the host plasma membrane during budding but is modified by viral envelope proteins. SU, the receptor binding protein, is hydrophilic and glycosylated and found on the surface of the particle. SU

is linked by disulphide bonds to TM, the transmembrane protein. TM mediates fusion of the viral and host membranes during entry into the cell [env proteins are reviewed by Hunter and Swanstrom (1990) and by Daar and Ho (1990)].

1.5 Life Cycle

1.5.1 Early Events

Early events are defined as those reactions that occur up to and including integration. These events depend upon virally-encoded proteins (figure 1.2).

1.5.1.1 Receptor Binding and Entry

The first event in viral infection is binding of SU to a specific host receptor. In the case of ecotropic MLVs, the host receptor has been cloned (Albritton, 1989) and identified as a cationic amino acid transporter (Kim *et al.*, 1991; Wang *et al.*, 1991). Once bound to the receptor, the viral envelope and cell membrane fuse and the virion core is released into the cytoplasm.

1.5.1.2 Reverse Transcription

Reverse transcription [reviewed by Goff (1990) and Whitcomb and Hughes (1992)] occurs in the cytoplasm and in association with the nucleocapsid proteins. Transcription

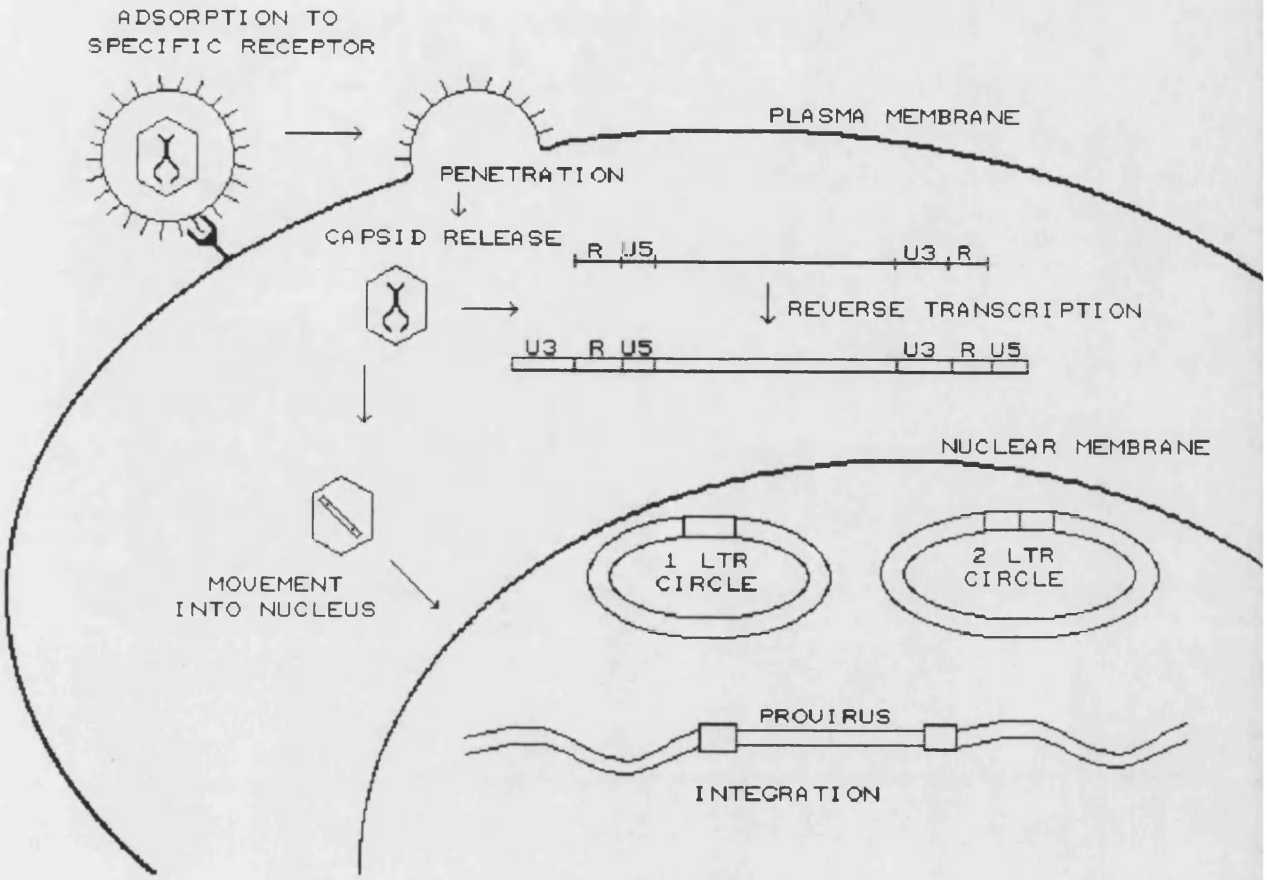


Figure 1.2 The Early Life Cycle. Following adsorption to the the receptor and penetration, the capsid is released into the cytoplasm where reverse transcription is completed. The viral DNA moves into the nucleus as a complex with viral proteins and is integrated into the host genome.

is primed by the host derived $tRNA^{Pro}$ binding to the pbs and proceeds in a 5' to 3' direction copying the 5' end of the genomic RNA. When the growing chain reaches the end of the template, the RNase H removes the hybridized RNA and the DNA-primer complex jumps to the 3' end of the genome, specifically to the R region. (The initial reverse transcript is called strong-stop DNA.) Strong-stop DNA transfer may either be inter or intra-strand (Hu and Temin, 1990). The DNA transcript then extends the full length of the genome while most of the RNA is removed by the RNase H. The first strand is now complete and is still attached to the tRNA. Second strand synthesis is primed by a polypurine tract, a fragment of the viral RNA. Synthesis continues through U3, R, and U5 and eventually encounters the tRNA which cannot be used as a template. The second jump then occurs, this time to DNA. This reaction is only intramolecular and obeys different rules from the first jump (Hu and Temin, 1990). Completion of the second strand and removal of the RNA primers complete the process.

Baltimore (1970) and Temin and Mizutani (1970) independently were the first to demonstrate an RNA-dependent DNA-polymerase activity associated with retrovirus particles. There is recent evidence to suggest that reverse transcription can occur in virions (Trono, 1992; Lori *et al.*, 1992) and that the reverse transcriptase may be active in its polyprotein precursor form (Hizi and Hughes, 1988; Peng *et al.*, 1991).

1.5.2 Late Events

Late events are defined as those occurring from transcription to budding and maturation. These events involve cellular enzymes (figure 1.3).

1.5.2.1 Transcription/Translation

The provirus is transcribed by cellular factors, producing a full-length RNA species that is capped and polyadenylated. How the level of mRNA expression is regulated is essentially unknown, but it seems to occur in similar proportion across the retroviruses. Some of the full-length mRNA is destined to be packaged in new virus particles although a proportion is spliced to form the *env* RNA.

Translation yields three polyprotein precursors: $Pr65^{gag}$, $Pr200^{gag\text{-}pol}$, and $Pr80^{env}$. Gag and gag-pol are translated from genomic-length RNA. Gag-pol expression is appoximately 5-10% of gag and in the case of MoMLV is read via an amber suppressor mutation and the insertion of a glutamine (Yoshinaka *et al.*, 1985). In other retroviruses there is a -1 frameshift that preserves the reading frame. A complex RNA pseudoknot structure mediates the read-through (Feng *et al.*, 1990; Wills *et al.*, 1991; Felsenstein and Goff, 1992). An 8 nucleotide purine-rich sequence 3' to the pseudoknot is also important (Feng *et al.*, 1992).

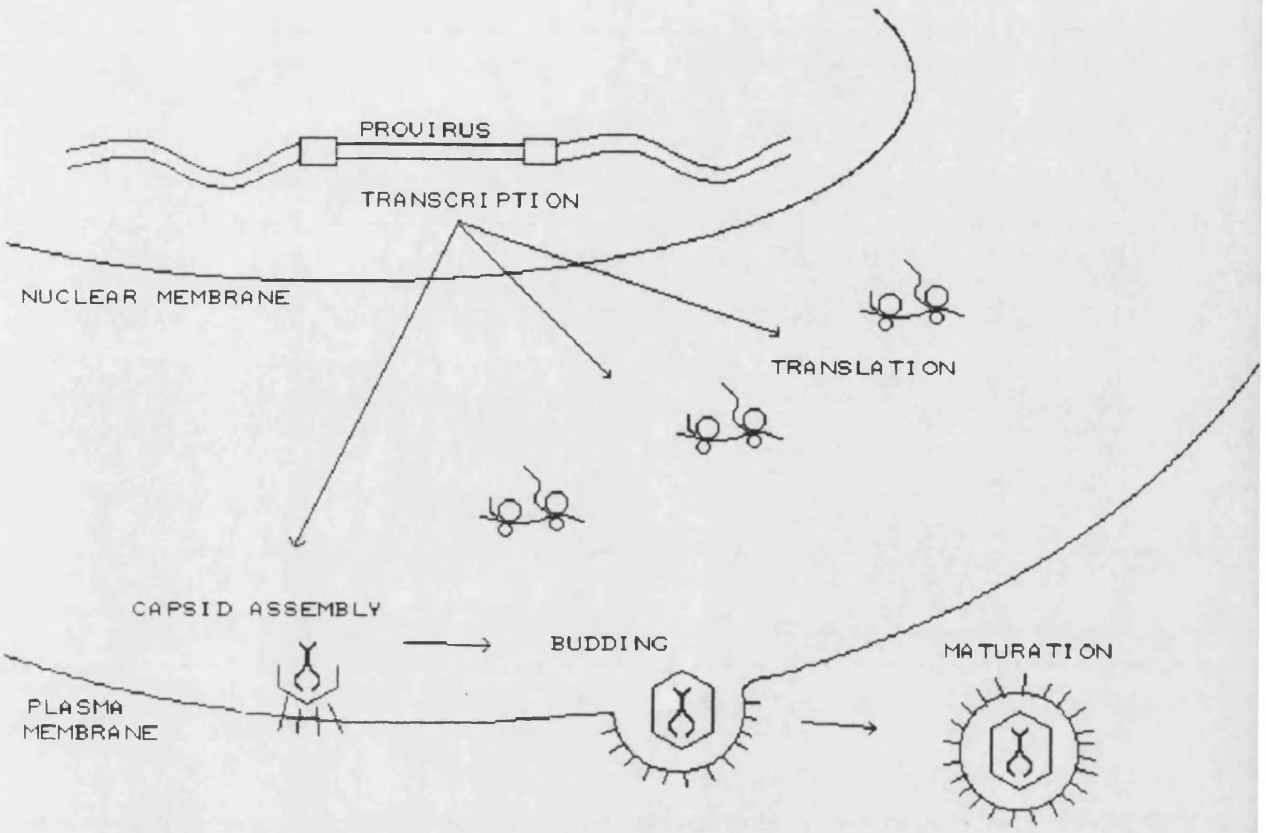


Figure 1.3 The Late Life Cycle. The provirus is transcribed and translated by cellular enzymes. Assembly takes place near the plasma membrane where some of the full-length mRNA is packaged into new virus particles. Following budding, the virion matures and becomes infectious.

1.5.2.2 Assembly, Budding, Maturation

Viral proteins and RNA assemble at the plasma membrane, brought together by hydrophobic groups at the N-terminus of the gag polyprotein. Once budding begins a crescent-shaped core is visible on electron micrographs and this closes into a circle once budding is complete. The virus is now in an immature, uninfective stage. Maturation is caused by the cleavage of the polyprotein precursors by the viral protease which itself is autocatalytically cleaved from gag-pol, accompanied by condensation of the core as seen on electron micrographs (Yoshinaka and Luftig, 1978). Condensation is associated with the disappearance of precursors and the appearance of env proteins on the surface (Yoshinaka and Luftig, 1977, 1978). Proteolytic cleavage is necessary to form the infective virus (Katoh *et al.*, 1985) (packaging is reviewed by Linial and Miller, 1990).

1.6 Integration

Integration is an essential step in the life cycle of most retroviruses (reviewed by Brown, 1990; Panganiban, 1990; Kulkosky and Skalka, 1990; Goff, 1992; Whitcomb and Hughes, 1992). SIV does not require integration for infectivity (Prakash *et al.*, 1992), while the integration requirement for HIV is less certain. Stevenson *et al.* (1990) report HIV antigen production continues in the absence of integration and that unintegrated viral DNA is transcriptionally active;

however Adachi *et al.* (1991) and Sakai *et al.* (1993) report the virus is uninfectious in the absence of integration. Although integration into the host genome occurs in other viruses (papova, adeno, hepadna and possibly herpes), it is by illegitimate recombination, not the specific, reproducible and precise mechanism by which retroviruses integrate. Once integrated, the provirus is maintained in dividing cells and is transmitted as an element of the host genome. A working model of the integration reaction has been developed from the results of *in vitro* assays. A summary and illustration of the integration reaction is presented in figure 1.4.

1.6.1 Structure of the Nucleoprotein Complex (NC)

The nucleoprotein complex consists of the viral DNA, integrase and possibly, in the case of MoMLV, CA protein (Bowerman *et al.*, 1989) and RT linked by disulphide and non-covalent interactions (Hu *et al.*, 1986). HIV-1 may differ in that only viral DNA and IN are detected with no evidence for *gag* gene products or RT (Farnet and Haseltine, 1991a). The structure of the unintegrated DNA is that expected from the process of reverse transcription; however, two nucleotides have been cleaved from the 3' ends (Fujiwara and Mizuuchi, 1988; Brown *et al.*, 1989). The nucleoprotein complex is capable of integration in the absence of cellular components and is necessary for integration as naked viral

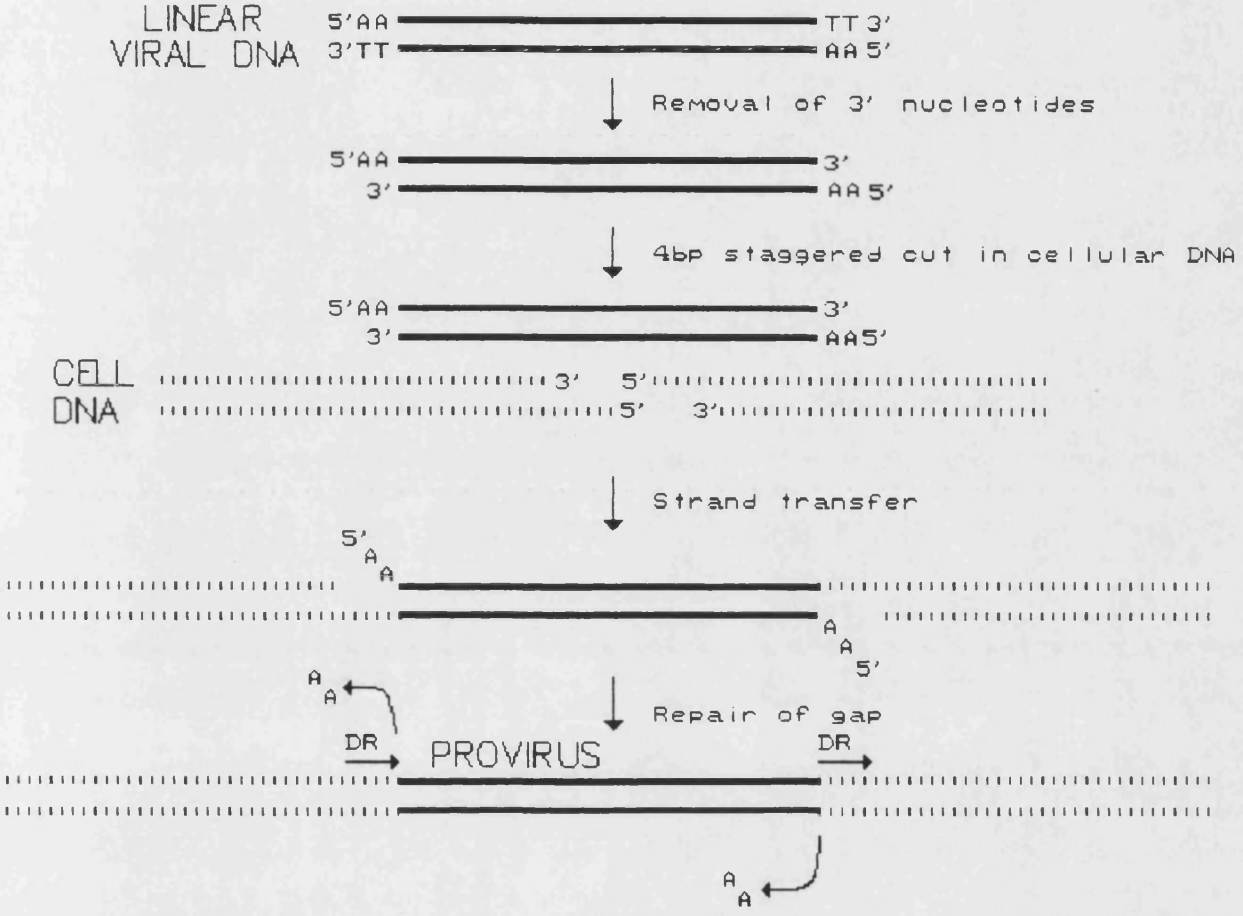


Figure 1.4 The Integration Reaction. The full-length viral DNA is processed by integrase removing 2 nucleotides from the 3' ends. The cellular DNA is cleaved in a 4 bp staggered cut and joined to the viral DNA. Repair of the gap follows the loss of two more nucleotides, leaving a 4 bp direct repeat (DR) and is accomplished by cellular enzymes.

DNA transfects poorly (Brown *et al.*, 1987; Bowerman *et al.*, 1989).

1.6.2 LTR Cleavage (3' Processing)

The viral LTR ends (*att* sites) are required in *cis* for integration. The *att* site is processed by the removal of two nucleotides from the 3' ends of the viral DNA (Dhar *et al.*, 1980; Brown *et al.*, 1989; Roth *et al.*, 1989; Ishimoto *et al.*, 1991). Initial reports suggested that only 1 bp was lost from the ends of HIV-1 DNA (Varmus and Brown, 1989), but subsequent studies established the loss of 2 bp (Pullen and Champoux, 1990; Pauza, 1990; Kulkosky *et al.*, 1990; Hong *et al.*, 1991). The endonuclease reaction is a site specific hydrolysis by nucleophilic attack of a phosphodiester bond. Alcohols can serve as the nucleophile for HIV (Engelman *et al.*,1991; Vink *et al.*, 1991a). There is no evidence for a covalent linkage for HIV but AMV and RSV do form a covalent linkage to the 5' ends of the DNA through serine or threonine residues (Katzman *et al.*, 1991). The two terminal 3' residues are removed as a dinucleotide (Fitzgerald *et al.*, 1991).

The viral ends form inverted repeats, which are a perfect palindrome in the case of MoMLV and imperfect in the case of HIV and avian viruses. Although the sequence of the inverted repeats varies widely across the retroviruses, each being virus specific, one feature is conserved: the sequence of the provirus is almost always 5'TG...CA3' (Varmus, 1982).

Many studies have defined the LTR sequence specificity and length of sequence required for correct integration (reviewed by Craigie, 1992). Colicelli and Goff (1985) and Basu and Varmus (1990) established the conserved CA sequence was more important than the identity of the terminal nucleotides. 3' processing for MoMLV is determined by the position of the CA, not by the distance from the LTR edge (Colicelli and Goff, 1988). HIV-1 IN will cleave up to 5 bp in mutant LTRs with additional nucleotides at the LTR ends (Vink *et al.*, 1991b). Cleavage appears to be directed by sequences internal to the LTR end. For MoMLV 7 bp are sufficient for cleavage (Murphy *et al.*, 1993) and the identity of bases 6 and 7 are most important for recognition (Bushman and Craigie, 1990). Avian viruses, with imperfect repeats at the LTR ends, have different requirements for each end. U5 requires a longer recognition sequence than U3 (Cobrinik *et al.*, 1987) with bases 3 and 4 most important for recognition in both (Katzman *et al.*, 1989; Cobrinik *et al.*, 1991). HIV, also with imperfect repeats, requires at least 9 bp (Bushman and Craigie, 1991; LaFemina *et al.*, 1991; Sherman *et al.*, 1992) with bases 3 and 6 important for recognition (Bushman and Craigie, 1992). The efficiency of 3' processing differs for the avian viruses with U3 greater than U5 for AMV (Vora *et al.*, 1990; Fitzgerald *et al.*, 1991) and U5 greater than U3 for HIV (Bushman and Craigie, 1991).

1.6.3 Entry into Nucleus

In order to gain access to the target DNA, the nucleoprotein complex must enter the nucleus. The process by which this occurs is poorly understood. Evidence from MoMLV suggests that mitosis, with the breakdown of the nuclear membrane, is required for integration (Roe *et al.*, 1993). This study established that integration does not occur when cells are arrested at G_1, S, and G_2, integration occurred only after metaphase. Li (1993) presents evidence that for HIV-1 integration only occurs during cell division. Evidence also suggests that the nucleoprotein complex may be actively transported to the nucleus independently of cell division via a nuclear localization signal in the MA protein (Bukrinsky et al 1993). However, there is no evidence that MA is part of the nucleoprotein complex (Farnet and Haseltine, 1991a). The possibility exists that the smaller complex of HIV-1 may be small enough to be transported through nuclear pores. One other observation, that of a nuclear localization signal identified in ALV IN, provides a third way of explaining entry (Lin and Grandgenett, 1991). RSV IN expressed in mammalian cells localizes to the nucleus (Mumm *et al.*, 1992). Clearly, further study is needed to resolve the question of how the nucleoprotein complex enters the nucleus; it may prove to be virus specific.

1.6.4 Integration Substrate

Once inside the nucleus the viral DNA takes one of three forms: linear, 2-LTR circles, and 1-LTR circles. The substrate for integration was originally thought to be the 2-LTR circle formed by the blunt-ended ligation of the ends of viral DNA, i.e. with the 3' termini intact (Panganiban and Temin, 1984; Grandgenett and Vora, 1985; Grandgenett *et al.*, 1986). The circle junction was subsequently shown not to be an efficient substrate for cleavage (Fujiwara and Mizuuchi, 1988; Brown *et al.*, 1989; Ellis and Bernstein, 1989; Lobel *et al.*, 1989) Thus it appears that the linear DNA is the immediate precursor to integration. In studies of HIV, 2 LTR circles were shown to result from blunt-ended ligation or auto-integration events with cellular factors required; 1 LTR circles result from recombination events (Farnet and Haseltine, 1991b). As yet, the role of the circular viral DNA is unknown.

1.6.5 Integration Site

There is little evidence to suggest sequence preferences at the site of proviral integration, despite the observation that many retroviruses preferentially integrate into specific chromosomal locations (reviewed by Craigie, 1992). Many studies have investigated the integration site specificity in both *in vitro* and *in vivo* systems. Pryciak and co-workers (1992a, 1992b) compared integration sites of

MoMLV in cell-free and *in vivo* assays, demonstrating viral DNA could integrate into either naked DNA or nucleosomal DNA but with a preference for nucleosomal DNA. The sequence preference was for AT rich regions, and a 10 bp periodic spacing suggested integration into the exposed face of the nucleosomal DNA helix. HIV and avian viruses also prefer AT rich regions (Hong, 1993; Fitzgerald *et al.*, 1992). There is also considerable evidence that MoMLV preferentially integrates into actively transcribing regions (Vijaya *et al.*, 1986; Rohdewohld *et al.*, 1987; Mooslehner *et al.*, 1990; Scherdin *et al.*, 1990)

1.6.6 Target Cleavage/Joining

The target DNA is cleaved in a 4 bp staggered cut for MoMLV (Shoemaker *et al.*, 1980; van Beveran, 1982), HIV has a 5 bp cut (Ellison et al 1990; Vincent et al 1990) while the avian retroviruses exhibit a 6 bp cleavage (Hughes *et al.*, 1981); cleavage was demonstrated to be IN specific not related to nucleotide sequence (Vink *et al.*, 1990). Following cleavage of the target DNA the viral 3' ends are joined to the 5' target ends. This reaction is by a single nucleophilic attack with transfer by a one-step transesterification (Engelman *et al.*, 1991) with no exogenous source of energy required (Brown *et al.*, 1990). Although AMV IN preferentially binds to AT rich regions, GC rich regions are prefered for the joining reaction (Grandgenett *et al.*,

1993). Also for AMV IN, the U3 strand transfer efficiency is greater than U5 (Fitzgerald *et al.*, 1992).

1.6.7 Resolution/Repair

Resolution of the integration intermediate is probably accomplished by cellular enzymes; there is no evidence to suggest that IN mediates resolution. Repair of the gap is accompanied by the loss of 2 nucleotides from the 5' viral DNA, followed by filling in of the missing nucleotides and leaving a characteristic direct repeat. The number of bases in the direct repeat is virus specific and is associated with the staggered cleavage; MoMLV leaves a 4 bp direct repeat, HIV 5 bp, and ALV/AMV 6 bp.

1.6.8 Disintegration

Disintegration is the reverse reaction of strand transfer. Given the integration intermediate as a substrate, IN catalyzes a cleavage of the LTR-host junction producing 3' recessed ends and ligates the target DNA. Both HIV-1 and MoMLV IN catalyze disintegration *in vitro* (Chow *et al.*, 1992; Jonsson *et al.*, 1993). The role of disintegration *in vivo* has yet to be determined.

1.7 Integrase

The retroviral integrase proteins share similar features, some of which are also shared by the retrotransposon integrases: the integrase coding region is located at the 3' portion of *pol*, all share a requirement for a divalent cation, and the mechanism of integration is similar. A comparison of 80 retroviral/retrotransposon IN sequences reveals a conserved N-terminal zinc finger motif and a central D,D(35)E region (Khan *et al.*, 1991). Characterization of the roles of these conserved domains has been primarily with HIV-1, HIV-2 and avian INs and can be generalized as follows: the zinc finger motif is involved in binding to the LTR, 3' processing, and strand transfer (Engelman and Craigie, 1992; van Gent *et al.*, 1992; Leavitt *et al.*, 1993; Vincent *et al.*, 1993) and the D,D(35)E region is involved in 3' processing, strand transfer, disintegration and functions as the catalytic centre (Khan *et al.*, 1991; Drelich *et al.*, 1992; Engelman and Craigie, 1992; Kulkosky *et al.*, 1992; LaFemina *et al.*, 1992; van Gent *et al.*, 1992; Bushman *et al.*, 1993; Leavitt *et al.*, 1993; van Gent *et al.*, 1993a, 1993b; Vink *et al.*, 1993). DNA binding is localized to the C-terminus of HIV-1 (Woerner and Marcus-Sekura, 1993) and avian IN (Mumm and Grandgenett, 1991). A leucine zipper motif has also been identified for RSV, HIV-1, and HIV-2 IN which may function as a dimerization signal (Lin and Grandgenett, 1991). A dimerization signal agrees with evidence that IN functions

as a multimer; avian IN as a dimer or tetramer (Grandgenett *et al.*, 1978; Katz *et al.*, 1992; Jones *et al.*, 1992) and HIV-1 as a dimer (Engelman *et al.*, 1993; Vincent *et al.*, 1993). Integrase can cross-react with the DNA of another virus, for instance HIV-1 and HIV-2 are interchangable (van Gent *et al.*, 1991) and MoMLV IN can act on HIV-1 DNA (Vink *et al.*, 1990).

1.7.1 MoMLV IN

MoMLV IN is a 46 kDa protein that is not phosphorylated (Tanese *et al.*, 1986). The divalent cation requirement is for Mn^{2+} (Craigie *et al.*, 1990). Whether or not Mg^{2+} can substitute has not been conclusively established. Craigie and co-workers (1990) report that Mg^{2+} can substitute *in vitro* in 3' processing but the reaction is less specific, cleaving 2 or 3 nucleotides from the 3' end. Jonsson *et al.* (1993) report no 3' processing or strand transfer with Mg^{2+} *in vitro*, while Ishimoto *et al.* (1991) report 3' processing in the presence of Mg^{2+} *in vitro* when the source of IN is from disrupted virions. Reaction conditions have been studied by Jonsson and co-workers (1993). The optimum Mn^{2+} concentration is 5 - 10 mM for 3' processing and strand transfer. KCl stimulates strand transfer, but NaCl inhibits the reaction. DTT is required for strand transfer, β-ME cannot substitute. For disintegration, the optimum Mn^{2+} concentration is 25 mM and β-ME can substitute for DTT.

1.7.2 HIV IN

HIV IN is a 32 kDa protein, is not phosphorylated and shares a 20% amino acid identity with MoMLV IN. Mn^{2+} is required for activity, but Ca^{2+} and Mg^{2+} can substitute (Billich *et al.*, 1992; Drelich *et al.*, 1992). Reaction conditions are different from those for MoMLV. KCl inhibits 3' processing and strand transfer and NaCl stimulates them (Craigie *et al.*, 1991; Engelman *et al.*, 1991; Vink *et al.*, 1991). DTT is required for strand transfer, but β-ME can substitute (Sherman and Fyfe, 1990).

1.7.3 Avian IN

Avian IN is a 32 kDa protein that is phosphorylated on a serine residue at the C-terminus (Horton *et al.*, 1988) but phosphorylation is not necessary for IN activity *in vitro* (Horton *et al.*, 1991). The processing of the pol proteins of avian retroviruses differs from MoMLV as RT is a heterodimer of α and β polypeptides. IN is derived from the C-terminal region of the β polypeptide with the C-terminal 4 kDa removed by the viral protease. The divalent cation requirement has not been determined with certainty. Mn^{2+} is required for 3' processing, strand transfer is reduced in the presence of Mg^{2+} (Katzman *et al.*, 1989) the specificity of 3' processing is different for Mn^{2+} or Mg^{2+}, with 2 nucleotides cleaved with Mg^{2+} and 2 or 3 with Mn^{2+} (Katz *et al.*, 1990). Fitzgerald *et al.* (1992) report that the cation

preference is dependent on the *in vitro* assay used. RSV IN is the only IN demonstrated to be capable of multiple turnover (Jones *et al.*, 1992).

1.7.4 Retrotransposons

Retroviral integrases appear to have an evolutionary relationship to the retrotransposons (Doolittle *et al.*, 1989; retrotransposons reviewed by Boeke, 1988; Boeke and Chapman, 1991). Similarities exist in genomic structure, organization and proteins, although in terms of the retrotransposon life cycle, lack an extracellular phase. The integration proteins have varying levels of sequence homology; conserved regions include the zinc finger motif and D,D(35)E region. The integration reaction is also similar. The integrase family (lambda integrase, FLP, and Cre) attach to broken DNA through a 3' phosphodiester linkage to tyrosine, and generate staggered breaks with 5' protruding ends of 6-8 bp. The resolvase and invertase family attach to broken DNA through a 5' phosphodiester linkage to a serine, generate staggered breaks with 3' protruding end of 2 bp. The integration mechanism of phage Mu most closely resembles the retroviral integration reaction (Mizuuchi and Craigie, 1986).

1.8 Aims of Work

With the world-wide pandemic of HIV infection, significant efforts are being made to discover anti-retroviral chemotherapeutics. Integrase is a unique target for antiviral chemotherapy as no cellular analogues of IN are known. A throrough understanding of the enzyme is needed in order to design specific inhibitors. At the outset of this project, MoMLV had been the subject of many more years of investigation than HIV. As a result, the IN protein from MoMLV was better characterized and understood than IN from HIV. MoMLV IN will serve as a model for the study of HIV IN and will complement work in this laboratory on the expression and characterization of HIV IN.

Therefore, the aims of this work are to develop an expression system for MoMLV IN in *E. coli* and demonstrate activity of the recombinant IN in *in vitro* assays. A kinetic assay for IN will be developed and then, using site-directed mutagenesis, the active and catalytic residues of IN will be investigated. The role of the zinc finger motif will also be investigated.

Chapter 2

This chapter describes the materials and methods that are common to Chapters 3-7.

2.1 Materials

2.1.1 Bacterial Strains, Plasmids and Culture Material

The following bacterial strains were used and their genotypes listed.

TG1- *supE hsd*Δ5 *thi*Δ(*lac-proAB*) F'[*traD*36 *proAB*$^+$ *lacI*^{q} *lacZ* M15].

TB1- *ara*Δ(*lac proAB*) *rspL*(ϕ80 *lacZ*ΔM15) *hsdR*.

JM105- *supE endA* sbc*B*15 *hsdR*4 *vpsL thi*Δ(*lac-proAB*) F'[*tra*D36 *proAB*$^+$ *lacI*^{q} *lacZ*ΔM15].

BL21(DE3)- *hsdS gal*(λ*c*I*ts*857 *ind*1 *S*am7 *nin*5 *lac*UV5-T7 gene*1*) CAG629 was used with permission from CellTech. The strain was made by Dr C. Cross (University of Wisconsin, USA). The genotype is unpublished but the parent strains are described in Baker *et al.* (1984).

Plasmids pTZ19R and pRIT5 were from Pharmacia. Plasmid pIF was the gift of Paul Krogstad and James Champoux (University of Washington-Seattle, USA). The construction of pIF is described in Krogstad and Champoux (1990). pMAL-c

and the purification system were from New England BioLabs. pRIT-5 was from Pharmacia.

Culture media were from Difco. LB medium was 1.0% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, 1.0% (w/v) NaCl. Storage plates for bacterial strains was LB agar (1.5% (w/v) bacto-agar in LB. Long-term storage of bacterial strains was in LB in 24% glycerol at -70°C. Ampicillin was from Sigma and used in a concentration of 0.1 mg/ml. Bacterial strain CAG629 was maintained and selected with tetracycline (Sigma) at a concentration of 40 µg/ml. *Hind*III and *Pst*I digested lambda DNA were from Pharmacia or Northumbria Biologicals.

2.1.2 Enzymes and Reagents

Goat anti-rabbit IgG alkaline phosphatase and horseradish peroxidase conjugates were from Sigma. [γ^{32}P]ATP and NENsorb columns were from NEN Dupont. Restriction enzymes were from Pharmacia, Northumbria Biologicals, or Boehringer-Mannheim. T4 polynucleotide kinase, IPTG, and Prepagene were from Northumbria Biologicals. GeneClean II was from Bio 101 Inc. Calf intestinal alkaline phosphatase and One-Phor-All buffer Plus were from Pharmacia. T4 DNA ligase was from Stratagene. Sequenase enzyme and sequencing kit was from United States Biochemical. Nitrocellulose and SDS-PAGE molecular weight markers were from BioRad. The ECL

detection system was from Amersham. Aristar methanol was from BDH. All other reagent were from BDH, Fisons or Sigma.

2.2 Methods

2.2.1 Horizontal Agarose Gel Electrophoresis

Horizontal agarose gels were used to analyze DNA. Gels were 1.0% agarose in 1x TAE (40 mM Tris-acetate, 1 mM EDTA). Ethidium bromide was added to a concentration of 0.5 µg/ml. The gels were run buffered in 1x TAE at 3-10 V/cm, lower voltage was used when the DNA was to be extracted from the gel. The loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 40% (w/v) sucrose) was added to the sample prior to loading the gel at a 0.2x volume. Molecular weight markers were *Hind*III or *Pst*I digested lambda DNA.

2.2.2 Preparation of Competent *E. coli*

E. coli was grown in LB to mid-log phase at 37°C with vigorous shaking. The culture was chilled on ice for 10 minutes, then centrifuged at 3000 rpm for 10 min. The cells were resuspended in ice-cold, sterile $CaCl_2$ solution (50 mM $CaCl_2$, 10 mM Tris-Cl (pH 8.0)) to one-half the original culture volume and placed on ice for 15 minutes. The suspension was centrifuged at 3,000 rpm for 10 minutes, the

supernatant discarded and the cells resuspended in $CaCl_2$ to one-fifteenth the original culture volume. The cell suspension was transferred to pre-chilled Eppendorf tubes in 200 µl aliquots and stored at 4°C for 16-24 hours.

2.2.3 Transformation of *E. coli*

Plasmid DNA, up to 10 ng, was added in 50 µl TE (10 mM Tris-Cl (pH 8.0), 1 mM EDTA (pH 8.0)) to competent *E. coli* cells. The bacterial suspension was placed on ice for 60 minutes, then transferred to a 42°C water bath for 2 minutes. LB (0.8 ml) was added and the tubes placed at 37°C for 1 hour to allow the antibiotic resistance gene to be expressed. LB agar plates with the appropriate antibiotic were inoculated with 250 µl of bacterial suspension, allowed to absorb and incubated at 37°C overnight.

2.2.4 Ethanol Precipitation of DNA

DNA was precipitated with 2.5 volumes cold absolute ethanol and one-tenth volume NaAc (3M NaAc, pH 4.8). The solution was placed at -20°C for 30-60 minutes or at -70°C for 15 minutes. The DNA was pelleted by centrifugation at 13,000 rpm for 10 minutes. The ethanol solution was removed, the pellet resuspended in 70% ethanol and vortexed. The DNA was pelleted again, the ethanol removed and the pellet dried in

a vacuum lyophilizer. The DNA was taken up in an appropriate volume of H_2O or TE.

2.2.5 Phenol/Chloroform Extraction of DNA

Phenol/chloroform extraction was used to purify DNA. An equal volume of phenol or phenol/chloroform (25:24:1 phenol:chloroform:iso-amyl-alcohol; phenol equilibrated to pH 8.0 with Tris-Cl, pH 8.0) was added to the DNA solution. The mixture was vortexed and centrifuged briefly to separate the aqueous and organic phases. The aqueous layer was transferred to a fresh Eppendorf tube and washed twice with chloroform (24:1 chloroform:iso-amyl-alcohol) to removed any residual phenol, vortexing and centrifuging as described above. The final aqueous phase was ethanol precipitated.

2.2.6 Rapid Preparation of Plasmid DNA (mini-prep)

Mini-preps were used for plasmid analysis only. 1.5-5 ml of an overnight culture of transformed bacteria were centrifuged at 13,000 rpm for 10 minutes. The medium was removed and the cells resuspended in 100 µl of ice-cold GTE (50 mM glucose, 10 mM EDTA, 25 mM Tris-Cl (pH 8.0)) and the cell suspension vortexed. After standing for 5 minutes at room temperature, 200 µl of 0.2 M NaOH, 1.0% (w/v) SDS was added and the tubes inverted 2-3 times, then stored on ice for 5 minutes. 150 µl of ice-cold KAc (3 M KAc (pH 4.8))

was added, vortexed gently, stored 5 minutes on ice and centrifuged at 13,000 rpm for 5 minutes. The supernatant was transferred to a fresh tube. The supernatent was then either ethanol precipitated, isopropanol precipitated (equal volume of isopropanol) or phenol/choroform extracted. Storage was in TE at -20°C.

2.2.7 Medium Scale Preparation of Plasmid DNA (mini-maxi prep)

Mini-maxi prep was done according to the method of Noguchi (1991). Briefly, *E. coli* transformed with the plasmid of choice was grown to saturation overnight in 50 ml LB with the appropriate antibiotic selection. The cells were pelleted by centrifugation at 3,000 rpm for 10 minutes then resuspended in 2 ml 50-1 TE (50 mM Tris-Cl (pH 8.0), 1 mM EDTA). The cells were pelleted by centrifugation at 13,000 rpm for 2 minutes and the pellet resuspended in 300 µl STE (15% (w/v) sucrose, 50 mM Tris-Cl (pH 8.0), 1 mM EDTA). Lysozyme (0.5 mg in STE) was added, the solution mixed by inversion, and briefly incubated on ice. TTE (300 µl; 0.1 % Triton X100, 50 mM Tris-Cl (pH 8.0), 50 mM EDTA) was added and the solution incubated a further 10 minutes on ice. DEPC (2 µl) was added, mixed by inversion, then boiled for 45 seconds. The cell debris was pelleted by centrifugation at 13,000 rpm for 10 minutes and the pellet removed. The supernatant was ethanol precipitated and the pellet

resuspended in water. The solution was treated with 50 µg RNase A at 37°C for 15 minutes then treated with 1 µl Proteinase K at 50°C for 15 minutes. NaCl was added to a final concentration of 0.2M and the solution extracted with phenol, then ethanol precipitated. Storage was in TE at -20°C.

2.2.8 Large Scale Preparation of Plasmid DNA (CsCl)

E. coli transformed with the plasmid of choice was grown to saturation in 200 ml LB. The cells were disrupted by alkaline lysis as described in 2.2.6 but with 40 times the volume of GTE, NaOH and KAc. The cell debris was pelleted by centrifugation at 3,000 rpm for 10 minutes and the supernatant transferred to a fresh tube. Cesium chloride was added to a concentration of 1.22 g/ml supernatant plus one-fifth volume of ethidium bromide (0.75 mg/ml). The solution spun overnight at 55,000 rpm in a Beckman T_i 50 rotor. The appropriate band was extracted under UV light and washed x3 with H_2O saturated butan-2-ol to remove the ethidium bromide. Two volumes of H_2O were added and the DNA ethanol precipitated. Storage was in TE at -20°C.

2.2.9 Large Scale Preparation of Plasmid DNA (LiCl)

E. coli transformed with the plasmid of choice was grown to saturation overnight in 100 ml LB with the appropriate

antibiotic selection. The cells were pelleted by centrifugation at 3000 rpm for 10 minutes. The medium was removed to fresh tubes and centrifuged again. The cells were lysed by the alkaline lysis method as described in 2.2.6. Volumes were: GTE, 4 ml; NaOH/SDS, 8 ml; and KAc, 6 ml. Following pelleting of the cell debris by centrifugation at 3000 rpm for 15 minutes, the supernatant was transferred to a fresh, sterile tube and mixed with 18 ml propan-2-ol. The solution was stored on ice 30 minutes and the nucleic acids pelleted by centrifugation at 3000 rpm for 30 minutes. The pellet was resuspended in 1 ml TE. LiCl was added to a final concentration of 3M, incubated on ice 30 minutes, then centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred to a fresh, sterile Eppendorf tube and ethanol precipitated. The DNA pellet was resuspended in 400 µl TE and treated with 40 µg RNase A at 37°C for 30 minutes. The reaction was stopped by the addition of 20 µl 10% SDS and heating to 75°C for 10 minutes. The sample was then LiCl precipitated again as before and the supernatant ethanol precipitated. The DNA was resuspended in TE, phenol extracted and ethanol precipitated. Storage was in water at -20°C.

2.2.10 Restriction Digestion

For analysis of plasmid DNA, reaction volumes contained 0.5-1.0 µg DNA, an excess of restriction enzyme and the

appropriate volume buffer according to the manufacturer's recommendation and sterile H_2O to a total volume of 10 µl. The reaction was carried out at 37°C for 1 hour and analyzed by agarose gel electrophoresis. Plasmids prepared by mini-prep were routinely treated with 10 µg RNase A during restriction digestion .

For preparation of DNA for cloning purposes, DNA was purified from CsCl or LiCl. Reactions contained 1-50 µg DNA, excess enzyme, and the appropriate amount of buffer. Reactions were carred out at 37°C for 2 hours or overnight. Double digestions were consecutive and each reaction stopped by heating to 85°C for 30 minutes and slow cooled. When buffer concentrations differed for each enzyme, the reaction requiring lower buffer concentration was carried out first. If buffers were incompatible, the samples were passaged through a Sepharose CL6B column and ethanol precipitated.

2.2.11 Oligonucleotide Purification

Oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesizer using phosphoramidite chemistry. The column containing freshly synthesized DNA was placed between two 2 ml syringes, one of which contained 1 ml ice-cold concentrated ammonia. Aliquots of 0.2 ml were drawn through the column at 20 minute intervals. A further 0.5 ml ammonia in a fresh syringe was passed through the column followed by 6 passages of the total volume between the two syringes.

The ammonia solution was placed in screw-top Eppendorfs and baked overnight at 55°C. After heating the solution was placed at -20°C for 10 minutes, the oligos were vacuum dried in a lyophilizer. The oligos were taken up in 300 µl H_2O and ethanol precipitated. Oligo sequences are listed in Table 2.1.

2.2.12 Quantitation of DNA

The concentration and purity of oligonucleotides were determined by OD_{260}/OD_{280} analysis. Concentration was based on 1 absorbance unit corresponding to approximately 37 µg of single-stranded DNA and 50 µg double-stranded DNA. An OD_{260}/OD_{280} ratio of greater than 1.8 was taken to indicate a pure sample.

2.2.13 End-labelling of DNA

Oligonucleotide length was determined by end labelling with [γ^{32}P]ATP using T4 polynucleotide kinase (PNK). Approximately 1-2 µg DNA was incubated with 7 U PNK, 10 µCi [γ^{32}P]ATP in kinase buffer (50 mM Tris-Cl (pH 8.0), 10 mM $MgCl_2$) in a total volume of 50 µl at 37°C for 30 minutes. The reaction was stopped with the addition of 10 µl F-dye (90% (v/v) formamide, 50 mM EDTA (pH 8.5), 0.1% (w/v) xylene cyanol and bromophenol blue). One tenth of the reaction volume was run on the appropriate percentage polyacrylamide

gel at 250V for 2-2.5 hours. Visualisation was by autoradiography.

2.2.14 Extraction of DNA from Agarose Gels

Three methods were used to purify DNA from agarose gels. Prepagene and GeneClean II were used according to the manufacturer's instructions. In the freeze-squeeze method the DNA was electrophoresed on an low melting point (LMP) agarose gel and the band of choice excised and placed in 0.5 ml Eppendorf tubes which had been pierced and packed with sterile, siliconised glass wool. The Eppendorf tube was frozen in liquid nitrogen for approximately 1-2 minutes, placed in a 1.5 ml Eppendorf and centrifuged at 13,000 rpm for 10 minutes. The solution in the 1.5 ml Eppendorf contained the DNA while the agarose remained in the 0.5 ml Eppendorf. DNA extracted by freeze squeeze was purified by phenol extraction or passage through a NENsorb column.

2.2.15 Purification of DNA on a NENsorb Column

The NENsorb column was first washed with 2 ml 100% methanol (Aristar), then equilibrated with 2 ml Reagent A (0.1 M Tris-Cl (pH 7.7), 10 mM triethylamine, 1 mM EDTA). The DNA sample was added to the column in TE or H_2O as well as a small amount of xylene cyanol and bromophenol blue. The column was washed with 2 ml Reagent A. Elution was with 1

ml 33% ethanol. The fraction that contained the DNA (which elutes with the xylene cyanol) was collected and ethanol precipitated.

2.2.16 Alkaline Phosphatase Treatment of DNA

Restriction digested plasmid DNA was dephosphorylated prior to ligations. Approximately 20-200 ng of vector DNA was incubated with 0.1 U calf intestinal alkaline phosphatase in One-Phor-All buffer at 37°C for 30 minutes. The reaction was stopped by heating to 85°C for 15 minutes.

2.2.17 Ligation of DNA

Ligations were carried out with a vector to insert ratio of 1:5 in 10 or 20 µl ligation buffer (50 mM Tris-Cl (pH 7.4), 10 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM ATP, 100 mg/ml BSA) with 2 U T_4 ligase at 15°C overnight. Approximately 50-200 ng of vector DNA was used. For blunt ended ligations the reaction was carried out in the presence of 1 µM hexaminecobalt chloride.

2.2.18 DNA Sequencing

DNA sequencing was done by the dideoxy chain-termination method of Sanger *et al.* (1977) using the Sequenase kit according to the manufacturer's instructions. Sequenase is

a modified T7 DNA polymerase. Plasmid DNA was denatured in 0.2 M NaOH, 0.2 mM EDTA at 37°C for 30 minutes then passaged through a CL6B spun column.

2.2.19 Protein Electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using the Mini-Atto apparatus. The resolving gel was 377mM Tris-Cl (pH 8.9), 7.5%, 8.0% or 10% acrylamide, (acrylamide:bis-acrylamide 37.5:1), 0.1% SDS and 0.625% (w/v) ammonium persulfate, 0.055% (v/v) TEMED. The solution was degassed prior to the addition of the SDS and ammonium persulfate. The stacking gel was 125 mM Tris-Cl (pH 6.75), 2.5% acrylamide, 0.1% SDS, and 0.3% (w/v) ammonium persulfate, 0.055% (v/v) TEMED, degassed as per the resolving gel. Protein samples were taken up in loading buffer (8M urea, 10 mM Tris-Cl (pH 8.0), 2 mM DTT, 0.2 mM DTPA, 0.1% SDS, 1% (v/v) β-mercaptoethanol, 0.05% (w/v) bromophenol blue) by boiling for 4-5 minutes. The gels were electrophoresed at 3 mA/cm in electrode buffer (0.125 M Tris, 1 M glycine, 0.5% (w/v) SDS). Staining of the gel was by Coomassie blue (0.25% (w/v) Coomassie Blue G250R, 45.5% (v/v) methanol) at room temperature for 30-45 minutes with agitation. Destaining (7.5% (v/v) glacial acetic acid, 5.0% (v/v) methanol) was carried out at room temperature or 37°C with agitation until the protein bands were clear and distinct.

2.2.20 Protein Transfer to Nitrocellulose

SDS-PAGE gels, nitrocellulose filter and filter papers were soaked in blotting transfer buffer (30 mM glycine, 48 mM Tris, 0.0375% (w/v) SDS, 20% methanol). The gel and filter were sandwiched between the appropriate thickness of filter papers. Transfer was performed in a semi-dry transfer blotter from Pharmacia at 0.8 mA/cm^2 for one hour 15 minutes.

2.2.21 Growth and Induction of Transformed Bacterial Strains

The expression of recombinant protein from *E. coli* was as follows. An overnight culture of *E. coli* transformed with the plasmid of choice was used to inoculate fresh medium in a volume ratio of 1:100; the appropriate antibiotic was included to maintain selection for the plasmid. Growth was usually at 32°C; growth at 37°C is noted where appropriate in the results sections. Cultures were grown to mid-log phase (OD_{590} 0.5-1.0) and induced with 1 mM IPTG. Growth was limited to 3 hours to reduce proteolytic degradation of the recombinant protein. At 3 hours, the cells were pelleted by centrifugation at 3000 rpm for 10 minutes and the medium removed. Cells were then frozen and stored for up to 1-2 months; or for preparation of whole cell lysate,

resuspended in 150 µl SDS-PAGE reducing buffer/ml original culture volume.

2.2.22 Sonication and Detergent Wash

E. coli were lysed by sonication prior to solubilization or purification of recombinant protein. Frozen cell paste was thawed and resuspended in 6 ml/g Basic Buffer (50 mM Tris-Cl (pH 8.0), 10 mM DTT, 1 mM DTPA). Sonication was done on ice at 50-150 W (depending on volume) x4 for 30 seconds at 1 minute intervals. Separation of insoluble from soluble material was accomplished by centrifugation at 13,000 rpm for 10 minutes. Insoluble fractions were washed in 4 ml/g original cell paste Wash Buffer (200mM NaCl, 0.5% (v/v) Nonidet P-40 in Basic Buffer). Insoluble material was pelleted by centrifugation at 13,000 rpm for 10 minutes.

2.2.23 Protein Determination

Protein concentrations were obtained using the Bradford assay. 900 µl Bradford reagent (0.01% (w/v) Coomassie Blue G-250, 5.0% ethanol, 8.5% (w/v) H_3PO_4) was added to 100 µl protein samples and the absorbance determined at 595 nm. Bovine serum albumin in concentrations of 0-1.0 mg/ml were used to produce protein calibration curves.

2.2.24 Immunoblotting

Following protein transfer the nitrocellulose filter was blocked in Blotto [5% Marvel non-fat dry milk, 0.02% sodium azide in PBST (0.01 % (v/v) Tween 20 in Phosphate Buffered Saline)] overnight at 4°C. After 2 washes in PBST the appropriate dilution of serum containing the primary antibody was added in Blotto and gently agitated for a minimum of 1 hour. Following 4 washes in PBST the primary antibody was labelled with goat anti-rabbit IgG alkaline phosphatase conjugate at the recommended dilution and agitated gently for 1 hour. The filter was washed twice with PBST and twice with TBST (150 mM NaCl, 50 mM Tris-Cl (pH 7.5), 0.01% (v/v) Tween 20). Detection was enzymatically with bromochloroindolyl phosphate/nitro blue tetrazolium substrate (BCIP/NBT) The filter was gently agitated in 0.33% BCIP/0.66% NBT (BCIP is 5% (w/v) BCIP in 100% DMF, NBT is 5% (w/v) NBT in 70% DMF) in Sigma Assay Buffer (ΣAB) (10mM glycine, 1 mM $MgCl_2$, 1 mM $ZnCl_2$, (pH 10.4)) until the bands were suitably dark. The reaction was stopped by washing in 2 changes of 20 mM EDTA (pH 8.0).

Pre-absorption of the primary antibody to remove anti-*E. coli* antibodies was done by incubating the primary antibody in PBST with a nitrocellulose filter transferred with 3x the usual amount of protein from whole cell lysate of the vector only control. Unless otherwise stated, primary anti-sera for all immunoblots were pre-absorbed.

Immunoblotting was also done using the ECL detection system according to the manufacturer's instructions. The second antibody was goat anti-rabbit IgG horseradish peroxidase conjugate.

Table 2.1 Sequence of oligonucleotides used for cloning, sequencing and activity assays. Note that oligo P5191 was used for both. All are listed 5'-3'.

CLONING

CysAla- ACTGAGACCGCCAAAGCTGCTGCACAAGTC

CysAla2- GACTTGTGCAGCAGCTTTGGCGGTCTCAGT

CysSer- ACTGAGACCTCCAAAGCTTCTGCACAAGTC

CysSer2- GACTTGTGCAGAAGCTTTGGAGGTCTCAGT

Eag- TCGAGCTCGGTACCCGGCCGGGGATCCATCCAGGGTAGGATAGAAAATTCA-TCACCCT

Eco- ACAGCTATGACCGATATCGAAAATTCATCA

Pst1- AGACTTGATGAGGACTGCAGCCGGGCGAAG

Pst2- CTGACCCCTCTGAGCTGCAGAAGAATTAAG

SEQUENCING

malE- GGTCGTCAGACTGTCGATGAAGCC

P4911- AACGCCAGCAAGTCTGC

P5191- CAGATCGGCCACTGTCTGAC

P5250- CAAAGCTCAGGCCAGGT

P5654- AGGACCTTTACACAGTCC

P5810- AATAAGGTTAACCCGCG

ASSAY

L1- CAGGTGGGGTCTTTCATT

L2C- AATGAAAGACCCCACCTG

Chapter 3

This chapter describes the synthesis of peptides corresponding to two regions of MoMLV IN and the production of immune sera.

3.1 Introduction

3.1.1 Solid Phase Peptide Synthesis (SPPS)

The development of fluorenylmethoxycarbonyl (Fmoc) polyamide SPPS (Atherton and Sheppard, 1985, 1987) allows peptides to be synthesized quickly and in quantity using automated synthesizers utilizing continuous monitoring systems. Peptides are synthesized from the C-terminus to the N-terminus, the C-terminus bound to a resin. The resin consists of a gelatinous polymer (acrylamide/bisacrylamide) within the pores of an inert support (Kieselguhr) that confers insolubility to the peptide chain. The resin is coupled to a linker prior to the addition of the first amino acid so that the resulting peptide will have an amide group at the C-terminus. The amino acids used in the synthesis are chemically modified in three ways. The amino group is protected by the base labile Fmoc, the carboxy group is modified by an activating group, and reactive side chains are protected by a variety of molecules.

Synthesis begins by activating the carboxy group to an active ester. The carboxy carbon of this ester undergoes

nucleophilic attack by the amino nitrogen of the resin bound peptide, forming a peptide bond. The Fmoc group is then removed and the next amino acid is introduced. Upon completion of the synthesis, the peptide must be cleaved from the resin and side chain protecting groups removed.

The monitoring system used in the peptide synthesis is counterion distribution monitoring (CDM)(Salisbury *et al.*, 1990). CDM utilizes an acidic dye and a base in solution with the incoming amino acids. The acidic dye forms ion pairs with any free base in solution or free base associated with the peptide chain. As the solution circulates through the column, the acidic dye initially binds to the basic sites on the growing peptide chain since there are more basic sites available on the support than in solution. As the amino acids react with amino groups of the peptide, dye is displaced and released into solution. When acylation is complete all the dye is in solution. The pre-column detector, set to the wavelength of the dye, measures the concentration of the dye as it circulates through the system.

3.1.2 Synthetic Peptides as Antigens

Synthetic peptides can elicit anti-peptide antibodies that cross react with the corresponding complete protein. This approach was used to produce antibodies reactive to the envelope protein of hepatitis B particles (Lerner, *et al.*, 1981) and to identify the proteolytic cleavage products of

the *pol* gene of yeast retrotransposon Ty1 (Garfinkel *et al.*, 1991). These antibodies can then be used in the isolation and purification of that protein. Although anti-peptide antibodies are reported to react extremely well in immunoblotting after the protein has been separated on a denaturing gel (van Regenmortel, 1988), it is recommended to select a sequence located on the surface of the molecule (Walter, 1986). In the absence of a crystal structure, evaluation of the hydrophilic and hydrophobic tendencies of residues along a polypeptide chain should identify regions likely to be found on the surface of a protein. Hopp and Woods (1981, 1983) found that the most hydrophilic regions of a protein corresponded to a continuous epitope. Other segments of a protein likely to be near the surface are highly flexible or mobile regions, shown by Geysen (1987) to have a higher than average antigenicity. Also, regions near the N- and C-terminus are usually found at the surface of a protein, are likely to be flexible and thus, antigenic. Immunization with free peptide, however, is unlikely to elicit an antibody response due to the small size of the peptide. Conjugation to a carrier protein raises the immunogenic potential of the peptide (Muller, 1988).

3.2. Materials and Methods

3.2.1 Materials

Fmoc amino acids and resin were from MilliGen. C_{18} RP-HPLC column was from Vydac. Dimethylformamide (DMF) was fractionally redistilled at 15 mmHg pressure and at 50°C. Hunter's TitreMax™ R-1 was from CytRx Corp. Goat anti-rabbit IgG alkaline phosphatase conjugate, para-nitro phenyl phosphate, thryglobulin, and m-Maleimidobonzoic acid N-Hydroxysuccinimide ester (MBS) were from Sigma. All reagents were Analar grade and were from either Sigma, Aldrich or BDH.

3.2.2 Methods

3.2.2.1 Computer Predictions of MoMLV IN Structure

The computer program Predict7 v1.01 (Cármenes *et al.*, 1988) was used to generate structural predictions of MoMLV IN. Numerical values were generated for hydrophilicity (Hopp and Woods, 1981), hydropathy (Kyte and Doolittle, 1982), flexibility (Karplus and Schulz, 1985), surface probability (Emini, 1985; Janin, 1978), and antigenicity (Welling, *et al.*, 1985). Secondary structure were classed as coil, turn, helix or sheet (Garnier *et al.*, 1978). Values were averaged over 6 residues, i.e. the value for residue 3 was averaged over residues 1 - 6.

3.2.2.2 Peptide Synthesis

Peptides were constructed on a MilliGen 9050 PepSynthesizer™ utilizing Fmoc polyamide solid phase chemistry. Table 2.1 lists the amino acids used for the synthesis.

A. Continuous UV Monitoring

Pre- and post-column detector channels allow continuous monitoring of the peptide assembly. Addition of the Fmoc-amino acid is indicated by a strong absorption followed by a series of oscillating peaks as the reagent circulates through the loop until uniform concentration is achieved. A rapid fall in absorption indicates the removal of excess reagent by washing. The deprotection step is monitored with the post-column detector which shows a deprotection peak as the Fmoc group is released.

B. Preparation of the Resin

Approximately 2 g of Pepsyn K (methyl ester resin) was incubated over night in ethylene diamine at room temperature to form

$$H_2N\text{-}CH_2\text{-}CH_2\text{-}NH\text{-}CO\text{-}[K]$$

where [K] is the resin.

Table 3.1 Fmoc Amino Acids Used in the Peptide Synthesis

Fmoc-L-Arg(Mtr)-OPfp
Fmoc-L-Asn-OPfp
Fmoc-L-Asp(OtBu)-OPfp
Fmoc-L-Cys(Trt)-OPfp
Fmoc-L-Gln-OPfp
Fmoc-L-Glu(OtBu)-OPfp
Fmoc-L-His(Boc)-OPfp
Fmoc-L-Ile-OPfp
Fmoc-L-Leu-OPfp
Fmoc-L-Lys(Boc)-OPfp
Fmoc-L-Pro-OPfp
Fmoc-L-Thr-(tBu)-ODhbt
Fmoc-L-Trp-(Boc)-OPfp
Fmoc-L-Trp-OPfp
Fmoc-L-Tyr(tBu)-OPfp
Fmoc-L-Val-OPfp

Abbreviations

Boc - *t*-Butoxycarbonyl
Fmoc - Fluorenylmethoxycarbonyl
Mtr - 4-Methoxy-2,3,6-trimethylbenzene sulphonyl
ODhbt - 3,4-Dihydro-4-oxo-benzotriazine-3-oxy
OPfp - Pentafluorophenoxy
OtBu - *t*-Butoxy
tBu - *t*-Butyl
Trt - Triphenylmethyl

C. Peptide Construction

The following steps were cycled by the synthesizer until the peptide was complete.

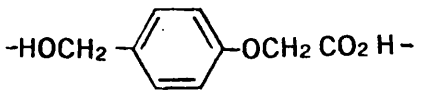
1. Deprotection: The column was flushed with 8 column volumes of 20% piperidine in DMF (dimethylformamide) to remove the amino protecting Fmoc group, until the absorbence at 365 nm returned to baseline.

2. DMF Wash: The piperidine was removed by washing with 12 column volumes of DMF.

3. Acylation: A 4-fold excess of amino acid dissolved in 0.33 M HOBt/DMF was cycled through the column until coupling was complete.

4. DMF Wash: Excess amino acids were washed away with 8 column volumes of DMF.

The first moiety to be added to the resin is the AM linker



which must be activated using diisopropylcarbodiimide (DIPCDI) to the hydroxybenzotriazole (HOBt) ester. DIPCDI

is added to the AM linker just prior to the acylation step. The result is an Fmoc capped AM linked resin.

D. Column Wash

Upon completion of the synthesis the resin was washed while still in the column to remove all traces of DMF. Approximately 25 ml of each reagent was passed through the column in the following sequence: t-amyl alcohol removes DMF, glacial acetic acid removes residual reagents, t-amyl alcohol removes glacial acetic acid, dichloromethane removes alcohol without shrinking the resin, diethyl ether shrinks the resin. Nitrogen was passed through the column to dry the resin. The resin was stored at 4°C until needed.

3.2.2.3 Trial Cleavage and Deprotection of Peptide KK28

To determine optimum cleavage and deprotection conditions, 50 mg of resin was incubated in 2 ml of a standard cleavage/deprotection mixture (90% (v/v) trifluoracetic acid (TFA) for cleavage, 2.5 % (v/v) H_2O and 7.5% (v/v) ethanedithiol as scavengers). Aliquots (300 µl) were taken at 1, 3, 5, and 18 hours. The samples were washed with 3 ml petroleum ether. The peptide was precipitated with 3 ml diethyl ether, centrifuged at 3,000 rpm for 5 min. and the pellets resuspended in 1 ml 5% B in A (B is 90% acetonitrile, 9.9% H_2O, 0.1% TFA, A is 0.1% TFA in H_2O).

The peptides were analyzed by analytical RP-HPLC on a C_{18} column.

3.2.2.4 Large Scale Preparation of KK28

Approximately 1 g of resin was incubated in 30 ml of the cleavage/deprotection mixture listed above for 1 hour at room temperature with occasional swirling. The resin was filtered off and the solution washed twice with an excess of petroleum ether. The peptide was precipitated with 30 ml diethyl ether, pelleted by centrifugation and resuspended in 30 ml 5% B in A. Samples were subjected to preparative RP-HPLC on a C_{18} column, the peak of interest collected, rotary evaporated and freeze-dried. Molecular weight was determined by mass spectroscopy using fast atom bombardment.

3.2.2.5 Trial Cleavage and Deprotection of Peptide KK334

To determine optimum conditions for the cleavage and deprotection of KK334, 50 mg of resin was incubated in 3 ml of cocktail A or cocktail K. Cocktail A is 95% (v/v) TFA for cleavage, 2.5% (w/v) phenol and 2.5% (v/v) ethanedithiol as scavengers. Cocktail K is 90% (v/v) TFA, 2.5% (w/v) phenol, 2.5% (v/v) ethanedithiol, 2.5% (v/v) H_2O, 2.5% (v/v) thioanisol as scavengers. Aliquots were taken at timed intervals and analyzed as described for the trial of KK28.

3.2.2.6 Large Scale Preparation of KK334

Approximately 1 g of the resin was incubated in Cocktail K for 17 hours at room temperature. Following the wash and precipitation, the pellet was resuspended in 50 ml 5% B in A. Preparation and analysis was as for KK28.

3.2.2.7 Peptide Coupling to Thyroglobulin

Peptides were coupled to thyroglobulin using m-Maleimidobenzoic acid N-Hydroxysuccinimide ester (MBS). In a total volume of 300 µl, 800 µg MBS (16 mg/ml in DMF) was reacted with 5 ng thyroglobulin (20 mg/ml in 10 mM Na_2PO_4 (pH 7.2)) for 30 minutes with gentle agitation. The solution was passed through a Sephadex G-25 column equilibrated with 50 mM Na_2PO_4 (pH 6.0) and eluted with the same buffer. Fractions were collected for 30 seconds each and analyzed by A_{280}. The first peak was pooled and reacted with 5 mg peptide (in 10 mM Na_2PO_4 (pH 6.0) through which nitrogen had been bubbled) for a minimum of 3 hours at room temperature with gentle agitation. Samples were stored at -20°C until needed.

3.2.2.8 Immunization Protocol

Samples of coupled peptide were thawed and brought to room temperature. Hunter's TitreMax adjuvant was also brought to room temperature and thoroughly vortexed. Approximately 0.3

mg (100 µl) was added to 200 µl Hunter's TitreMax and emulsified by passage between 2 1 ml all plastic syringes. A further 100 µl peptide was added to the emulsion and passaged between the syringes until a change in viscosity was detected (i.e. thickening). The emulsion was used promptly. During the emulsification a 50% loss of volume was observed. Two sandy half-lop rabbits were immunised with each peptide, 50 µl in each hind flank. Pre-immune sera were collected before immunization. Test bleeds were collected at intervals described in the Results. Test bleeds (~5 ml) were allowed to clot overnight at 4°C, the serum taken off and spun twice at 13,000 rpm to remove any red blood cells. Large volumes of blood (>50 ml) were allowed to clot at 4°C for 6-8 hours, centrifuged at 2,400 rpm for 10 minutes, the serum transferred to fresh tubes and centrifuged at 2,400 rpm for 10 minutes. The slower centrifugation speed decreased the incidence of haemolysis. Sera were stored at -20°C.

3.2.2.9 Direct Enzyme Linked Immunoabsorbant Assay (ELISA)

Microtitre plates were coated with 100 µg antigen in 50 mM $NaCO_3$ (pH 9.5) for 2 hours at 4°C. The antigen solution was removed and the plates blocked with Blotto overnight at 4°C. Following 4 washes with PBST, the primary antibody (rabbit serum) was added in a 1:100 dilution in PBST. Serial dilutions were carried out to 1:3200 and the plates incubated for 2 hours at 37°C. The plates were washed x4 in

PBST and the appropriate dilution goat anti-rabbit IgG alkaline phosphatase conjugate added in PBST. Incubation was at 37°C for 1 hour. The plates were washed twice in PBST and twice in Sigma Assay Buffer (ΣAB). Detection was enzymatically using a 1 mg/ml solution of para-nitro phenyl phosphate in ΣAB. The reaction proceeded until sufficient colour was seen, then stopped with EDTA (pH 8.0) to a final concentration of 50 mM. Absorbance was read at 405 nm.

3.3 Results

3.3.1 Structural Predictions of MoMLV IN

Two putative antigenic regions of MoMLV IN were identified by computer predictions of secondary structure. Region 1 near the N-terminus encompassed residues 28-42. Analysis showed high values for hydrophilicity, flexibility, antigenicity, and a high probability of surface location, especially for residues 30-35 (Figure 3.1). Region 2, residues 334-348, showed a high surface probability over the entire region as well as the highest hydrophilicity values over the entire protein (Figure 3.2) Regions 1 and 2 were chosen for peptide synthesis for production of anti-sera.

3.3.2 Peptide Synthesis

3.3.2.1 Synthesis of Peptide KK28

KK28 corresponded to Region 1 described above with the sequence

H_2N-CIYDKTKKYWVYQGKP-COOH

The N-terminal non-coding cysteine was placed for the purpose of coupling to the carrier protein. Coupling efficiencies for the first 14 amino acids (residues 16-3) were all above 99%. Ile-2 gave a coupling of 98.2%, the final Cys-1, 96%. Double coupling reactions were not performed. Coupling efficiencies matched detector readings, deprotection peaks were sharp, large and uniform throughout.

3.3.2.2 Synthesis of Peptide KK334

KK334 corresponded to Region 2 described above with the sequence

H_2N-CWVRRHQTKNLEPRWK-COOH

The N-terminal non-coding cysteine was again placed for the purpose of coupling to the carrier protein. Fmoc-Trp(Boc)-OPfp was utilized to minimize the modification of Trp residues by the Mtr protecting group of Arg during

deprotection (Sieber, 1987; Riniker and Hartman, 1990; White, 1992). Coupling efficiencies for the first 12 amino acids were all above 98.9%. Arg-13 fell to 96.4% and Arg-14 to 92.7%. A double coupling was performed for Arg 14 repeating the acylation step. Coupling efficiency then rose to 98%. Val-15 and Trp-16 were scored by the computer as having 101.6 % efficiencies, but this was contradicted in the lower absorbence reading by the column detector. Trp-16 underwent double coupling with a reported 97.8% efficiency, Cys-17 99.5%. The much lower values reported by the column detector did not agree with coupling efficiency and this pointed to a calibration defect in either the synthesizer or the software, or to inherent difficulties in the CDM system. Deprotection peaks were large, sharp and uniform throughout.

3.3.3 Peptide Purification

3.3.3.1 Purification of KK28

Small scale cleavage and deprotection trials were analyzed by RP-HPLC on a 5-60% acetonitrile/water gradient. A single large peak was identified eluting at 30% in each sample (1, 3, 5, and 18 hours). Extending the reaction time past one hour did not significantly improve the yield of the putative full length product. Large scale purification was performed by preparative RP-HPLC on a 5-60% acetonitrile gradient stepped over 20-30%. Fractions were collected and analyzed.

Pure fractions were pooled and prepared as described in 3.2.2.4.

3.3.3.2 Purification of KK334

Aliquots of the trial cleavage and deprotection were taken at 5.5, 22, 29 and 48 hours of cocktails A and K. RP-HPLC analysis over a 5-60% acetonitrile gradient showed a major peak eluting at 29% acetonitrile in both cocktails A and K in the 5.5 hour sample. By 22 hours, several other peaks were noted, a major one at 36%, in both cocktails. Extending the time to 29 and 48 hours resulted in an increase in the number and size of peaks eluting in the 30-45% acetonitrile range. Cocktail K consistently showed a larger peak at 29% and fewer smaller peaks. Trials were repeated using cocktail K and taking aliquots at 17 and 22 hours. RP-HPLC analysis showed 2 major peaks at 29% and 36%. Since the peak height ratio (29%:36%) was greater at 17 hours, it was decided to purify 100 mg resin, cleave and deprotect for 17 hours, purify the 29% peak and analyze the peptide by mass spectroscopy. Large scale purification as described in 3.2.2.6 was undertaken following confirmation of a peptide the correct size.

3.3.4 Mass Spectroscopy

Purified peptides were analyzed by mass spectroscopy by fast atom bombardment (SERC Mass Spectroscopy Service Centre,

University College, Swansea). The molecular weight of KK28 reported was 2090. This did not correspond to the expected molecular weight of 2019. A check of the sequence desired and the sequence synthesized revealed a Gln in position 14 that should have been a Gly. With Gln in position 14, the molecular weight of 2090 agrees perfectly. The expected molecular weight of KK334 was 2136, the experimental result was 2137.

3.3.5 Production of Anti-Peptide Anti-Sera

3.3.5.1 Immune Response of Rabbits 88 and 89 to KK28

Rabbits 88 and 89 were immunised with peptide KK28 as described in Methods. The first bleed 21 days post-immunization was analyzed by ELISA. Rabbit 88 showed a strong response to thyroglobulin and a weaker response to KK28. Rabbit 89 showed a poor response to thyroglobulin as compared to pre-immune serum and no response to KK28 (data not shown). Both rabbits were boosted with the same quantity of antigen and test bleeds obtained 12 and 28 days post-boost. Rabbit 89 showed no response at all to KK28 and was released from the experiment. Rabbit 88's titre to KK28 peaked with the second bleed as shown in Figure 3.5. A second boost was performed with the same quantity of antigen as initally and bled 9 days post-boost. ELISA showed a continuing fall in titre. A fifth bleed was obtained 26 days post-second boost with a very slight rise in titre to

KK28. A third boost was performed using 0.1 g of coupled peptide in 1 ml PBS, 0.5 ml injected into each hind flank. A sixth test bleed was obtained 9 days post-boost and tested by ELISA. Again, a very slight rise was noted in titre levels (data not shown). Being unable to raise the titre level to it's previous high level, the rabbit was exsanguinated while under anaesthesia and the serum stored.

3.3.5.2 Immune Response of Rabbits 73 and 80 to KK334

Having learned from Rabbit 88 that titre levels could not be maintained, the protocol was changed for Rabbits 73 and 80. Following initial inoculation with KK334, a first test at 21 days detected very weak responses in both animals. A boost was performed and test bleeds obtained 9 days post-boost. ELISA showed titre levels in Rabbit 80 comparable to that of Rabbit 88's highest titre (Figure 3.6). Rabbit 73 showed no immune response to KK334, despite high titre to thyroglobulin. Rabbit 73 was released from the experiment and Rabbit 80 exsanguinated while under anaesthesia. Peptide specificity was checked by probing peptide KK28 with anti-KK334 and *vice versa*. The results showed no cross-reactivity and demonstrated that the anti-sera recognized a specific peptide.

3.3.6 Immunoblotting with Anti-peptide Anti-sera

TG1 transformed with pTZ19R or pIF were prepared as described in Chapter 4. Aliquots of the insoluble fractions after the wash step were separated on an SDS-PAGE gel and transferred to nitrocellulose. Western blotting was as described in Methods. Both anti-KK28 and anti-KK334 sera detect a major band of approximately 43 kDa that is not seen in the vector only controls (Figure 3.7) and co-migrates with the IN band identified on Coomassie stained SDS-PAGE gels.

Pre-immune serum, test bleed sera and the final serum from Rabbits 88 (anti-KK28) and Rabbit 80 (anti-KK334) were used to probe whole cell lysate from TG1 transformed with either PTZ19R or pIF (Figure 3.8). Anti-KK28 again recognizes the 43 kDa protein not seen in pTZ19R lysate in bleeds 2 through the final bleed. This band is not recognized in the pre-immune or first bleed sera. A unique band cannot be identified in the pIF lysates probed with anti-KK334. The discrepancy between Figure 3.7B and 3.8B is discussed in section 3.4.4.

Figure 3.1 Graphical representation of the output of Predict7 for residues 18-52 of IN. Values for antigenicity were multiplied by a factor of 10. Values for flexibility were modified by subtracting 1 and multiplying by 10. Thes manipulations were for the sake of clarity. The region to particularly note is between residues 30-35 with high value for all parameters. Residues 28-42 were chosen for synthesis.

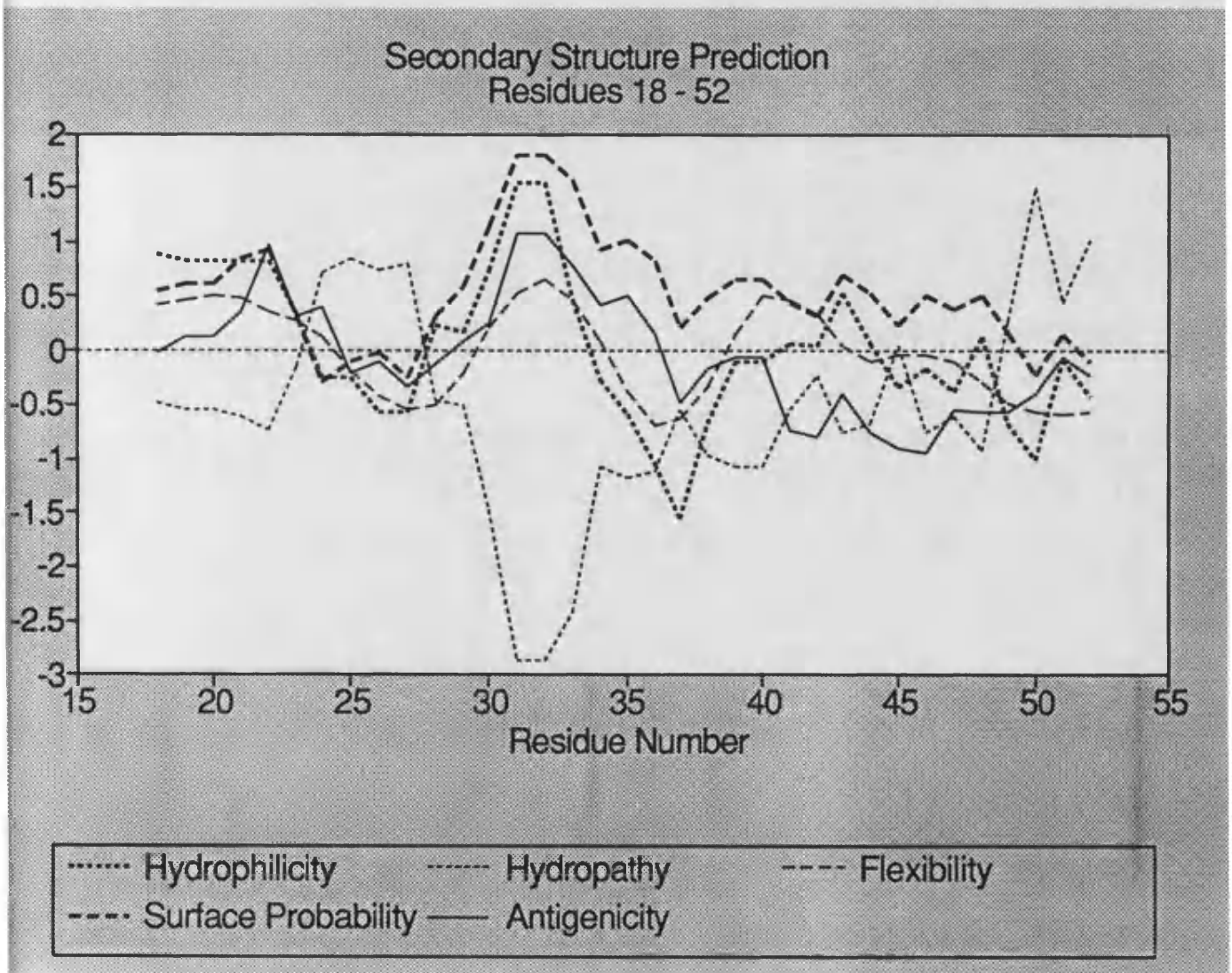

Secondary Structure Prediction
Residues 18 - 52
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
15
20
25
30
35
40
45
50
55
Residue Number
Hydrophilicity
Hydropathy
Flexibility
Surface Probability
Antigenicity

Figure 3.2 Graphical representation of the output of Predict7 for residues 324-358 of IN. Values for antigenicity and flexibility were altered as described for figure 3.1 for the sake of clarity. Residues 337-342 are noted for the highest hydropathy values over the entire protein. Residues 334-348 were chosen for synthesis.

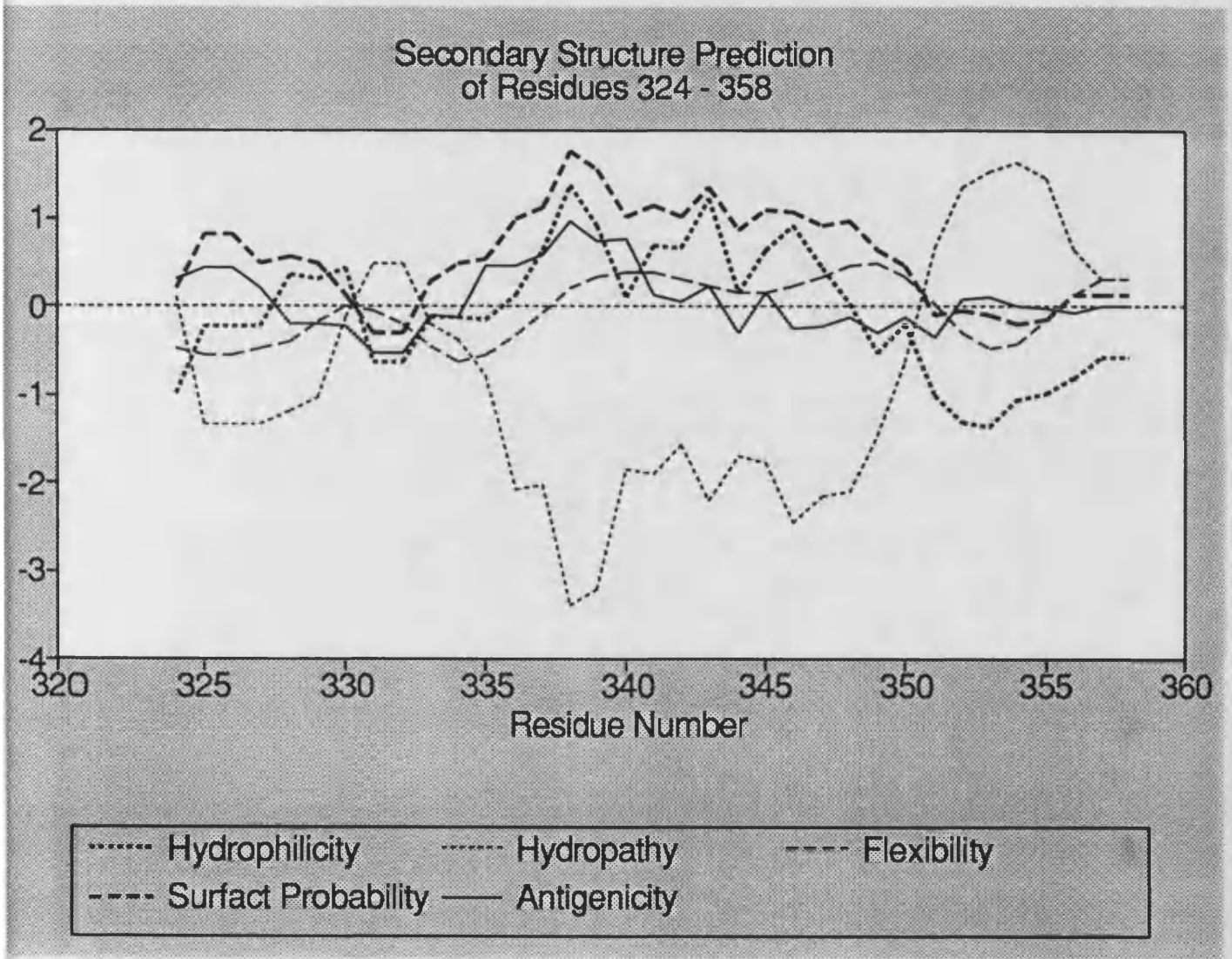

Secondary Structure Prediction
of Residues 324 - 358
2
1
0
-1
-2
-3
-4
320
325
330
335
340
345
350
355
360
Residue Number
Hydrophilicity
Hydropathy
Flexibility
Surfact Probability
Antigenicity

Figure 3.3 Bar graph depicting coupling efficiencies during the synthesis of KK28. The residues are in the order of synthesis, C-terminus to N-terminus. Efficiencies above 100% reflect amino acid derivatives with a similar colour t the dye.

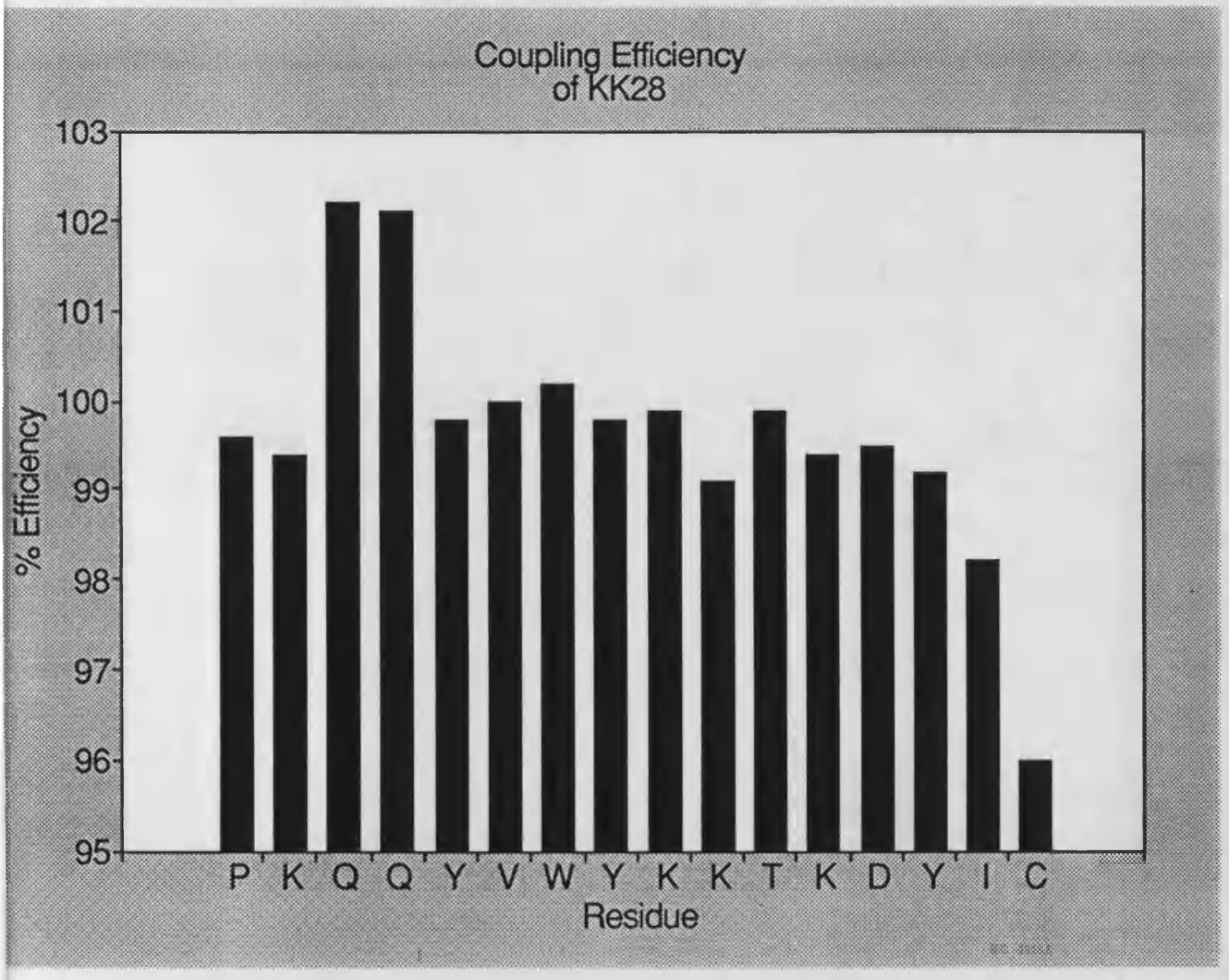

Coupling Efficiency
of KK28
% Efficiency
103
102
101
100
99
98
97
96
95
P K Q Q Y V W Y K K T K D Y I C
Residue

Figure 3.4 Bar graph depicting coupling efficiencies during the synthesis of KK334. Starred residues reflect double couplings. Efficiencies above 100% for the T and Q residue reflect derivataves with a similar colour to the dye. Coupling efficiencies of the final 5 residues are discussed in sections 3.3.3.2 and 3.4.1.

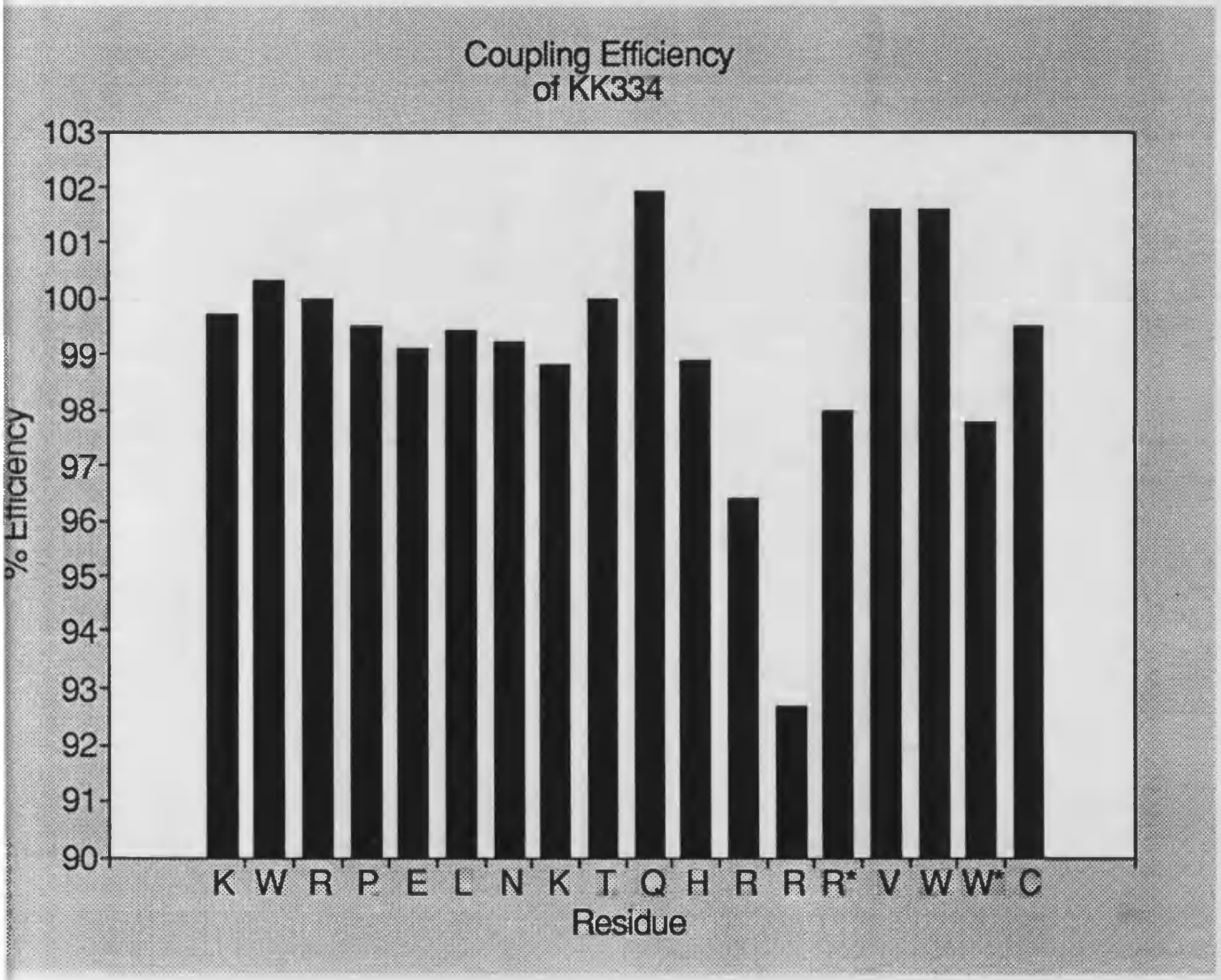

Coupling Efficiency
of KK334
% Efficiency
103
102
101
100
99
98
97
96
95
94
93
92
91
90
K W R P E L N K T Q H R R R* V W W* C
Residue

Figure 3.5 ELISA of Rabbit 88 against peptide KK28. Microtitre plates were prepared with 100 µg peptide. Goat anti-rabbit IgG alkaline phosphatase conjugate was a 1:2500 dilution. Detection was enzymatically with para-nitro phenyl phosphate. The ELISA shows a peak titre with the 2nd bleed, 12 days post-1st boost.

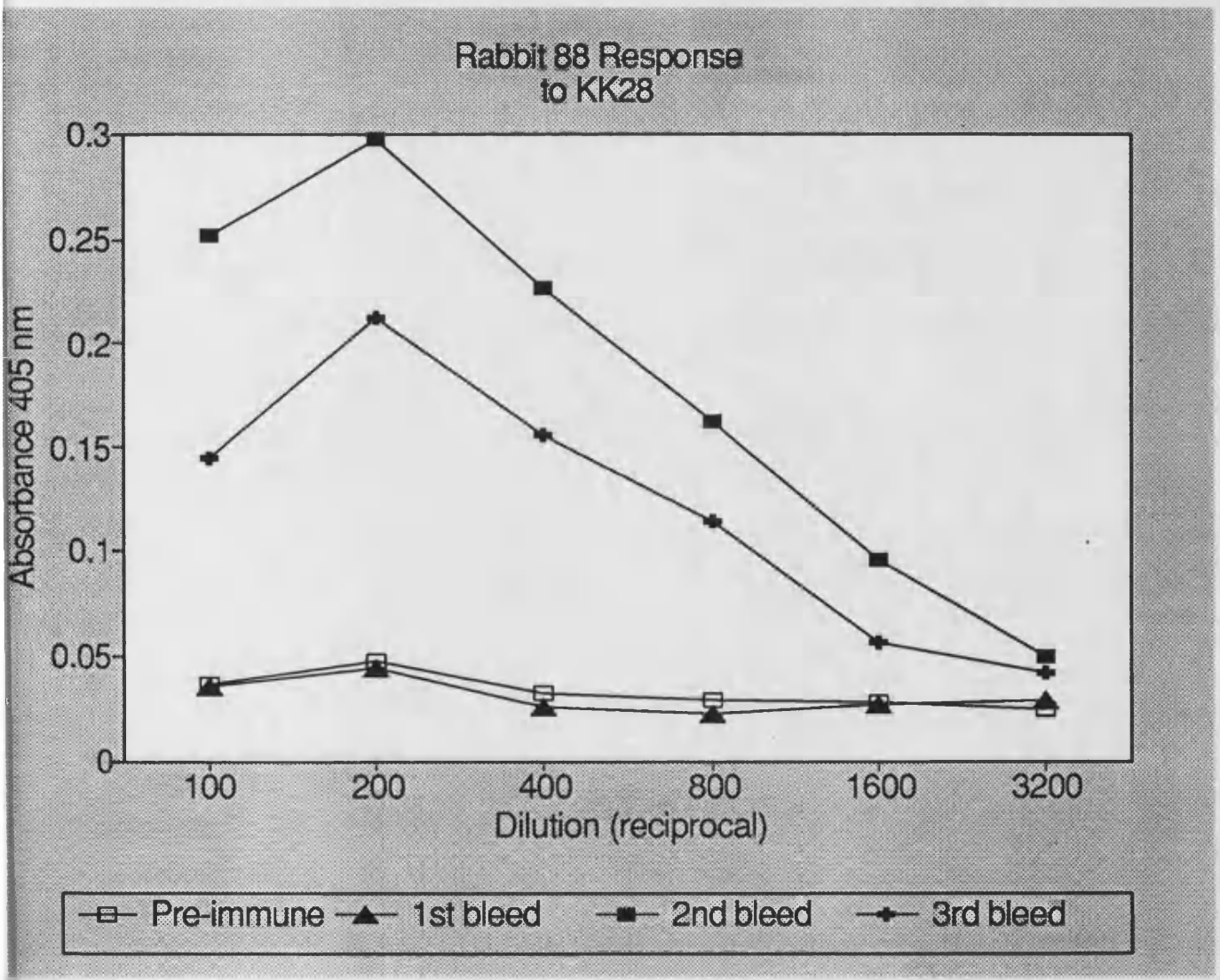

Rabbit 88 Response
to KK28
Absorbance 405 nm
0.3
0.25
0.2
0.15
0.1
0.05
0
100
200
400
800
1600
3200
Dilution (reciprocal)
Pre-immune
1st bleed
2nd bleed
3rd bleed

Figure 3.6 ELISA of Rabbit 80 against peptide KK334. Microtitre plates were prepared with 100 µg peptide. Goat anti-rabbit IgG alkaline phosphatase conjugate was a 1:5000 dilution. Detection was enzymatically with para-nitro phenyl phosphate. The ELISA shows a peak titre with the 2nd bleed, 9 days post-1st boost.

Pre-peptide=pre-immune serum against KK334
Pre-thyro=pre-immune serum against thyroglobulin
2nd-peptide=2nd bleed serum against KK334
2nd-thyro=2nd bleed serum against thyroglobulin

First bleeds are not shown.

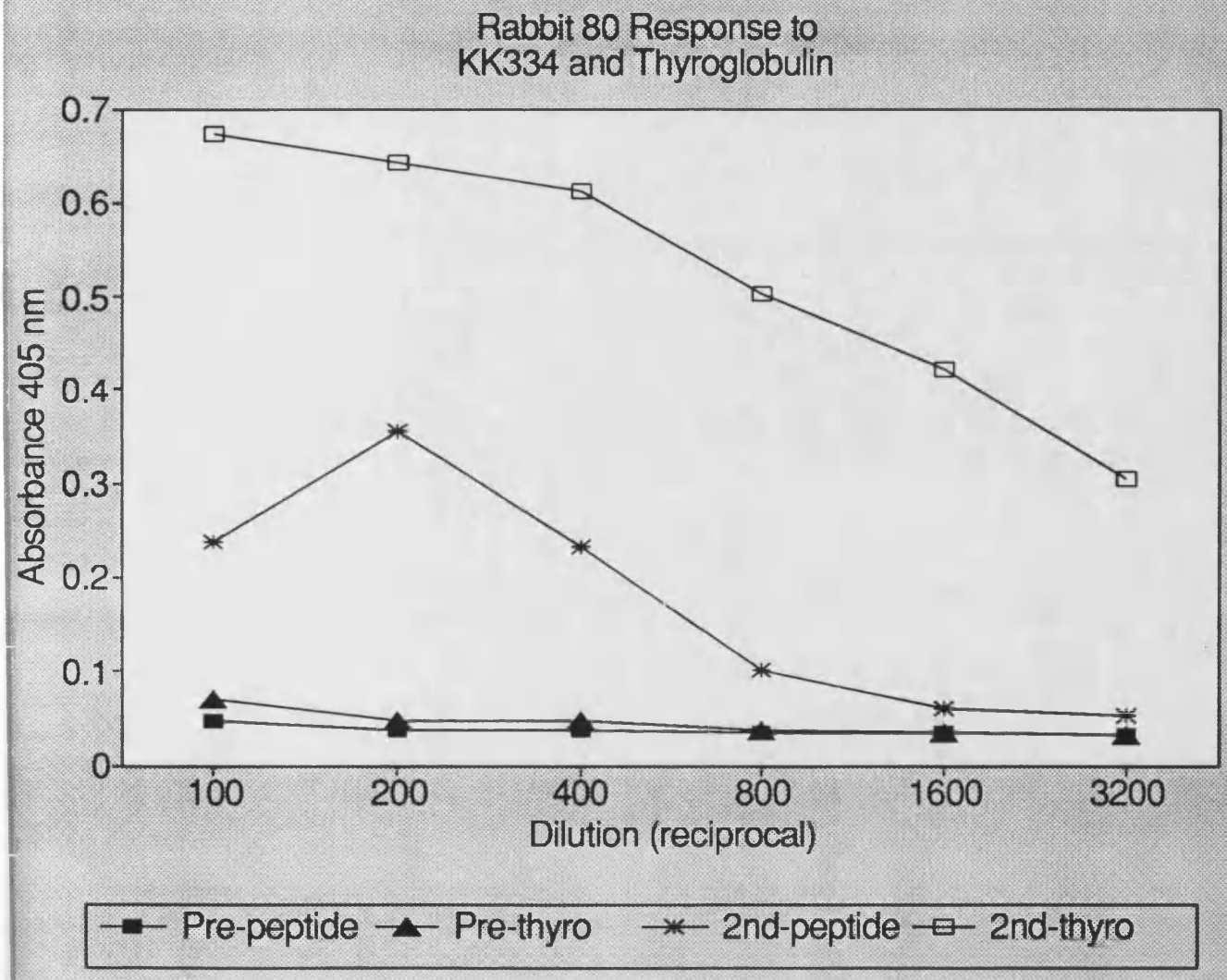

Rabbit 80 Response to
KK334 and Thyroglobulin
Absorbance 405 nm
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
200
400
800
1600
3200
Dilution (reciprocal)
Pre-peptide
Pre-thyro
2nd-peptide
2nd-thyro

Figure 3.7 Immunoblot with anti-KK28 and anti-KK334 antisera. *E. coli* TG1 transformed with either pTZ19R or pIF were grown, induced, sonicated and washed as described in section 2.2.22. Insoluble pellets were resuspended in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel and transferred to nitrocellulose. Filters were probed with a 1:400 dilution anti-KK28 or a 1:100 dilution of anti-KK334. Goat anti-rabbit IgG alkalin phosphatase conjugate was a 1:5000 dilution. Visualizatio was with NBT/BCIP.

A. Anti-KK28

B. Anti-KK334

1. TG1 transformed with pTZ19R

2. TG1 transformed with pIF

The arrows indicate the position of IN.

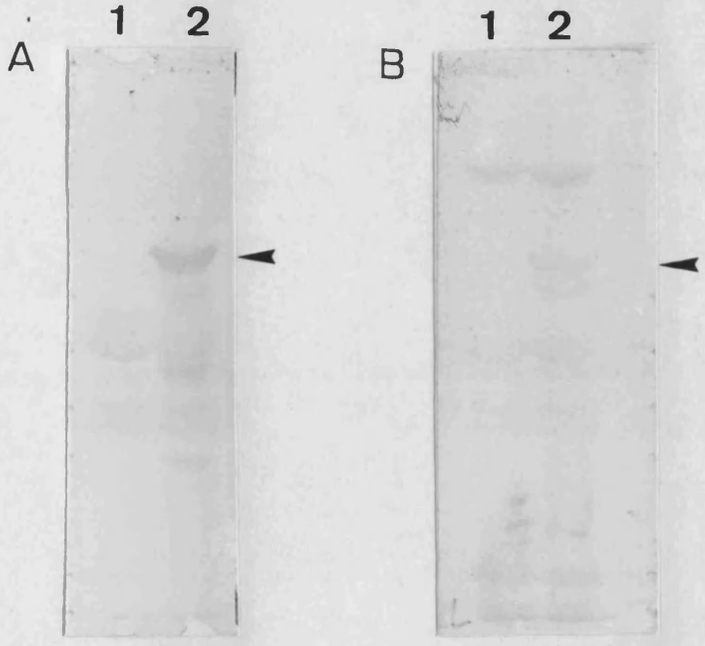

1 2
A
1 2
B

Figure 3.8 Immunoblots of all sera collected. Whole cell lysates of TG1 cells transformed with either pTZ19R or pIF were resuspended in reducing SDS-PAGE sample buffer, separated by a 10% polyacrylamide gel, and transferred to nitrocellulose. Anti-sera were a 1:100 dilution. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualization was with NBT/BCIP.

A. KK28 anti-serum

1 - pre-immune

2 - 1st bleed

3 - 2nd bleed

4 - 3rd bleed

5 - 4th bleed

6 - 5th bleed

7 - 6th bleed

8 - final bleed

B. KK334 anti-serum

1 - pre-immune

2 - 1st bleed

3 - 2nd bleed

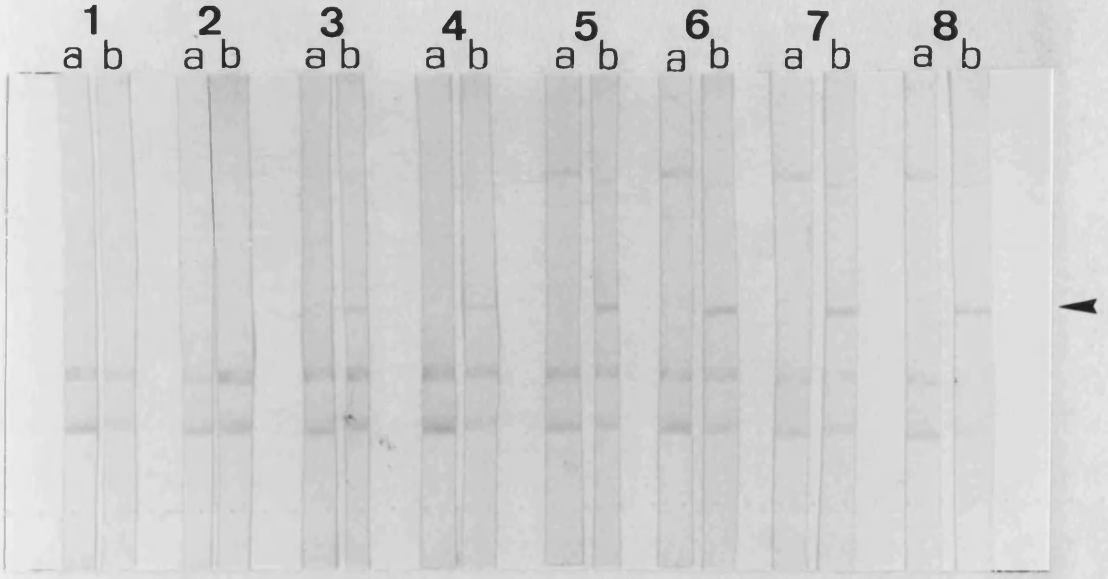
4 - final bleed

a. TG1 transformed with pTZ19R

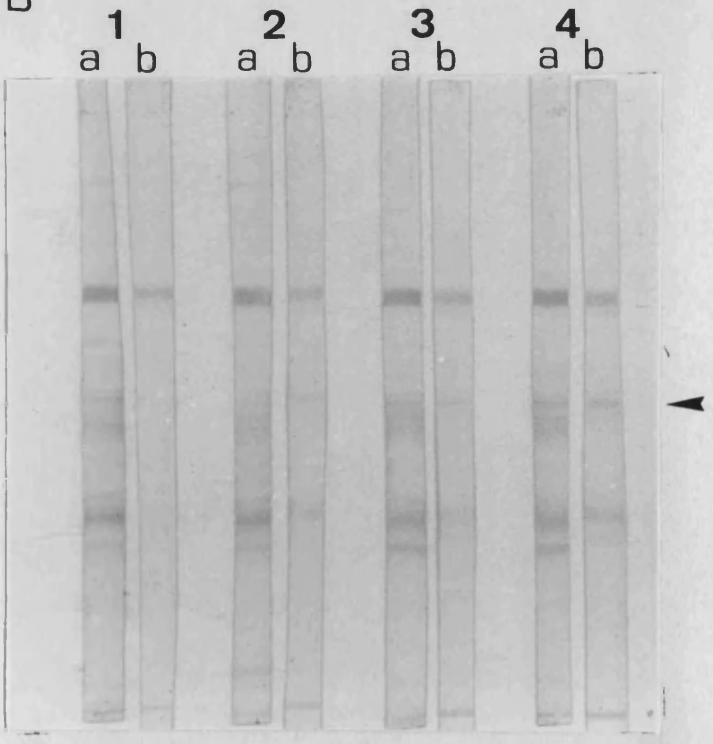
b. TG1 transformed with pIF

Arrows indicate the position of IN.

A



B



3.4 Discussion

3.4.1 Peptide Coupling Efficiency

During the synthesis of both KK28 and KK334 the column detectors registered low coupling efficiencies. In other words, indications were that dye had been removed from the system, the dye was presumably still bound to free amino acids in the peptide and these free amino acids were not coupled to the incoming amino acids. The linear nature of SPPS requires that coupling be 99% or greater for each coupling. At 99% efficiency in the synthesis of a 16-mer one would expect $.99^{16}$ x 100% purity, or 85%. An intuitive answer to why coupling efficiencies might fall during a synthesis is the formation of secondary structure as the peptide chain is lengthened. Westall and Robinson (1970) encountered this problem and included urea in the reaction to destabilize secondary structure, resulting in an increased yield of full-length product. This answer is not fully satisfactory when using the CDM system. The dye molecule used (Quinoline Yellow) is bulky, far larger than the incoming amino acid. If secondary structure were casusing steric hindrance to the incoming amino acid, then the dye molecule would be sterically hindered as well and ought to give a false high coupling reading. Non-ionic dye interactions with the resin and the peptide chain cause false low readings (Kinsman and Olivier, in press). The low coupling readings may instead be a byproduct of CDM and not

true incomplete couplings. The observation has also been made that that most acylations following Arg are reported low; this is likely to be an interaction between the dye and the previous Arg (R. Kinsman, personal communication). CDM is efficient at measuring the rate of acylation but less so the extent (Kinsman and Olivier, in press). The only true test of detecting incomplete acylation and the resulting deletion peptide would be to purify and characterize the deletion product. Both KK28 and KK334 were purified as full-length product as shown by the mass spectroscopy data.

3.4.2 The Sequence of KK28

Peptide KK28 was synthesized with a Gln in place of Gly in position 14. The change results in the substitution of a conformationally mobile residue to a bulkier, polar residue. ELISA demonstrated that anti-KK28 was able to recognize peptide KK28, but more importantly, immunoblotting demonstrated that KK28 anti-serum was able to recognize IN, despite the non-native peptide. Sutcliff and co-workers (1983) demonstrated that peptides 6 residues or less in length fail to produce immune serum capable of recognizing full-length protein. Peptides 8-13 residues were sufficient to produce an immune response, suggesting that the shortest sequence recognizable to an antibody is 8 residues. KK28 was 15 residues in length, with the predicted highest antigenic region at residues 2-7. Residue 14 may not form part of the antibody binding site. A method to test this

hypothesis would be to synthesize a shorter peptide, 10-12 residues in length, deleting residues 13-16, produce anti-peptide anti-serum and test the ability of this anti-serum to recognize IN.

3.4.3 Production of Anti-peptide Anti-sera

Peptide-carrier coupling efficiency was not measured. Deen and co-workers, (1990) found that "psuedo-coupling", non-covalent interactions between the peptide and carrier protein, interfered with chemical methods for determining coupling efficiency. The pseudo conjugates elicited no antibody response. Based on these observations, determination of coupling efficiency was not necessary.

An alternative to complete Freund's adjuvant (CFA) was sought as CFA is known to cause significant toxic effects such as local, subpleural, hepatic and renal granuloma formation and necrotizing dermatitis (Warren *et al.*, 1986; Johnston *et al.*, 1991; Leskowitz and Waksman, 1960; Broderson, 1989). Hunter's TitreMax™ R-1 is one new alternative to CFA, reported by the manufacturer to cause a minimum of side effects. The manufacturer's instructions for use in rabbits is minimal. Boosting is said to be rarely necessary. Very little response to peptide was detected by ELISA after the initial innoculation in all rabbits. Response was greater after one boost. In the case of KK28 and Rabbit 88 the titre peaked at the second bleed and subsequent boosts were unable to raise the titre. Smith

et al. (1992) in a comparison of CFA and TitreMax also found boosting to be necessary, although in his experiment, the TitreMax titre continued to rise after only one boost. The immunoblots of all the anti-sera to KK28 showed the anti-sera recongized IN at bleeds 2-final demonstrating that the lowest titre, bleed 3, was sufficient to recognize IN at a dilution of 1:100. It should be noted that TitreMax was less toxic to the rabbits. Observations by the animal house staff noted no lesions for Rabbits 73, 80 and 89. Rabbit 88 developed a small nodule on the left hind flank, noted at the time of the final bleed. Rabbit 88 had been injected 4 times, 3 with TitreMax, once with antigen in PBS.

3.4.4 KK334 Anti-sera

The immunobots in figures 3.7B and 3.8B show a discrepancy in the ability of anti-KK334 to recognize IN. There are two differences between those blots. The protein sample in figure 3.7B was the insoluble pellet after sonication and detergent wash and the serum was pre-absorbed to remove *E. coli* antibodies. In figure 3.8B the proteins were whole cell lysate, the serum had not been pre-absorbed. These results demonstrate that for anti-KK334 pre-absorption is a necessary step and that the anti-serum recognized IN more efficiently after some *E. coli* proteins had been removed. Due to these observations, anti-KK334 was rarely utilized.

3.5 Conclusions

The amino acid sequence of MoMLV IN was analyzed for secondary structure and two regions chosen for peptide synthesis. The peptides were synthesized, purified, and molecular weight determined by mass spectroscopy. Anti-sera was produced in rabbits using Hunter's TitreMax™ R-1 as the adjuvant with a minimum of side effects to the rabbits. The anti-sera were capable of recognizing recombinant IN on immunoblots, although KK28 was more efficient than KK334.

Chapter 4

Chapter 4 describes the expression and solubilization of IN as expressed as a fusion with the first three amino acids of β-galactosidase.

4.1 Introduction

Studies of integration have utilized four sources of IN:

i) IN purified from virus particles (Colicelli and Goff, 1985, 1988b; Panet and Baltimore, 1987; Brown *et al.*, 1987, 1989; Luk *et al.*, 1987; Fujiwara and Mizuuchi, 1988; Fujiwara and Craigie, 1989; Roth *et al.*, 1990; Ishimoto *et al.*, 1991; Roth, 1991)

ii) expression in insect cells using recombinant baculovirus (Craigie *et al.*, 1990; Bushman *et al.*, 1990)

iii) expression in *E. coli* (Terry *et al.*, 1988; Hizi and Hughes, 1988; Sherman and Fyfe, 1990; Krogstad and Champoux, 1990; Marcus-Sekura *et al.*, 1990; Vink *et al.*, 1991b; Marczinovitz *et al.*, 1992; Bushman *et al.*, 1993; Roth *et al.*, 1990)

iv) expression in yeast (Leavitt *et al.*, 1993; Basu and Varmus 1990).

IN-associated activites can be demonstrated *in vitro* using purified IN from virions and from baculovirus expression; however, expression in *E. coli* has proven problematic. IN as expressed from *E. coli* usually aggregates into insoluble inclusion bodies necessitating denaturation and refolding.

That proper refolding is not always successful is inferred from numerous reports in which only non-specific DNA binding is demonstrated (for example, Luk *et al.*, 1987; Hizi and Hughes, 1988; Krogstad and Champoux, 1990; Roth *et al.*, 1990).

Inclusion bodies are aggregates of recombinant protein, the predominant component, with bacterial proteins such as the subunits of RNA polymerase, outer membrane proteins, 16S and 23S rRNA and plasmid DNA (Hartley and Kane, 1988). Why inclusion bodies form is not well understood. Formation does not seem to be a response to foreign proteins, as *E. coli* proteins expressed to high levels also form inclusion bodies (Cheng, 1983; Gribskov and Burgess, 1983, Botterman and Zabeau, 1985). Nor is aggregation associated with high levels of recombinant protein, as low levels can also be insoluble (Schoemaker et al 1985). Various strategies can be employed to purify, denature, and refold recombinant proteins and are reviewed by Marston (1986), Jaenicke (1991), and Fischer *et al.* (1993). Although some denatured proteins can refold in the absence of other proteins *in vitro*, *in vivo* folding involves accessory proteins, such as protein disulphide isomerase and peptidyl-prolyl-*cis-trans*-isomerase [reviewed by Gething and Sambrook (1992), Schmid (1993), and Jaenicke, (1993)]. Polypeptides with incorrect prolyl bonds can only partially fold which promotes aggregation and may be the cause of some proteins being aggregated into insoluble inclusion bodies.

4.2 Methods

4.2.1 Solubilization of Proteins with Urea and Guanidine

Insoluble pellets following detergent wash (section 2.2.22) were resuspended in 4 ml urea or guanidine (in 50 mM Tris-Cl, pH 8.0) per g original cell paste in concentrations of 2, 4, 6 or 8 M. Samples were allowed to stand overnight at room temperature then centrifuged at 13,000 rpm for 10 minutes to separate soluble from insoluble material. Soluble fractions were either dialyzed against various buffers as described in figure 4.3 or frozen at -20°C. Guanidine soluble fractions were routinely TCA precipitated before SDS-PAGE analysis.

4.2.2 TCA Precipitation of Proteins

Proteins were precipitated in 10% trichloroacetic acid, stored on ice 15 minutes and washed x3 in 50% diethyl ether, 50% ethanol. The pellet was allowed to dry completely before resuspending in reducing SDS-PAGE sample buffer.

4.3 Results

4.3.1 Expression of IN

In virions, IN is synthesized as a polyprotein precursor which is cleaved by the viral protease to form the mature

IN. Expression in *E. coli* necessitates providing a Met or expression as a fusion. Plasmid pIF constructed by Krogstad and Champoux (1990) expresses IN as a fusion with three amino acids from β-galactosidase. Plasmids pIF and pTZ19R, the parent vector of pIF, were transformed into *E. coli* TG1 by the $CaCl_2$ method and recombinants selected by ampicillin resistance. Plasmid was prepared for analysis by mini-prep, restriction digested and analyzed by agarose gel electrophoresis. An insert of 1.2 kb was confirmed. Whole cell lysates were prepared and analyzed by SDS-PAGE stained with Coomasie Blue. An additional protein was seen in the pIF construct that was not seen in the vector-only control (figure 4.1). Its molecular weight was calculated from a calibration curve to be 45 kDa (figure 4.2). Four major degradation products were identified (figure 4.7 lane 10). The molecular weights were calculated and are 39.6, 32.5, 29.7, and 25.7 kDa (figure 4.2).

4.3.2 Solubilization of IN

Figure 4.3 lists the steps taken to produce soluble IN. It should be noted that all these trials were analyzed by Coomassie-stained SDS-PAGE gels. The presence of a bacterial protein that co-migrated with the IN and displayed similar solubility characteristics greatly complicated the identification of IN in these gels. Figure 4.4 illustrates this problem and shows that following sonication and washing the IN is located in the insoluble fraction. That IN was

being expressed in insoluble inclusion bodies was confirmed by light microscopy of intact bacterial cells. The inclusion bodies appear as dark spheres at the ends of the cells. Initial denaturation and refolding trials were conducted with growth at 37°C and included all the conditions listed in figure 4.3. Optimum denaturation time and temperature were investigated and found to be at room temperature left standing overnight. The IN was visible on gels only in the insoluble fractions in the various attempts to renature the protein, even when the proteins were concentrated by TCA precipitation, which indicates reaggregation. No significant difference between Buffers A and R was observed.

Two conditions of inclusion body formation were investigated. As inclusion bodies can be associated with high levels of expression, induction was limited to 1 hour, after which the cells were sonicated and analyzed by SDS-PAGE. No IN was visible on Coomassie stained gels. A lower growth temperature has been suggested to decrease the incidence of protein aggregation (Schein and Noteborn, 1988; Schein, 1989). The effect of lowered growth temperature was investigated on the protocols listed in figure 4.3 with the exception of the dilution experiments and direct dialysis into SDS and buffers A and R. The results showed a greater proportion of IN soluble in 6M guanidine. All subsequent bacterial growth was done at the lower temperature of 30°C.

When the anti-sera were available, a number of these experiments were repeated and analyzed by immunoblotting,

presented in figures 4.5-4.7. The results confirmed that *E. coli* expressed IN was insoluble in urea, but soluble in 6M guanidine, and did reveal a proportion of IN soluble after step-wise dialysis against decreasing concentrations of urea. An aliquot of the soluble portion was tested for activity, presented in Chapter 6.

4.3.3 Expression in *E. coli* CAG629 and Bl21

To investigate the effects of expression of IN in other bacterial strains, plasmids pTZ19R and pIF were tranformed into BL21, a strain successfully used in the expression of HIV IN (Sherman and Fyfe, 1991) and CAG629, a protease deficient strain. As BL21 is a lysogenic strain of bacteria, it is necessary to determine optimum induction conditions. Analysis of an induction trial is presented in figure 4.8 with sonicate and wash soluble fractions; very low levels of expression were observed, with some soluble IN after sonication, but none in the wash soluble fraction. With CAG629 solubilizations were carred out through urea. Analysis by Coomassie-stained SDS-PAGE gels and immunoblotting revealed low levels of expression, with IN only visualized with difficulty. Analysis by immunoblotting with the more sensitive ECL detection system revealed IN to be soluble in all concentrations of urea (figure 4.9).

Figure 4.1 Coomassie stained SDS-PAGE gel of protein expression in *E. coli* TG1. Whole cell lysates were resuspended in reducing SDS-PAGE sample buffer and loaded a 7.5% acrylamide gel. Staining was with Coomassie Blue. The arrow indicates the position of IN.

Lane 1 Molecular weight standards (kDa)

Lane 2 TG1 transformed with pTZ19R

Lane 3 TG1 transformed with pIF

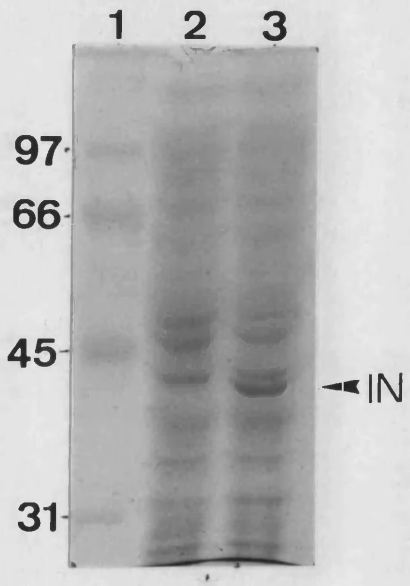

1
2
3
97
66
45
31
IN

Figure 4.2 Calibration curve of the molecular weight of IN expressed in TG1; A-D are the major degradation products identified from figure 4.7 lane 10.

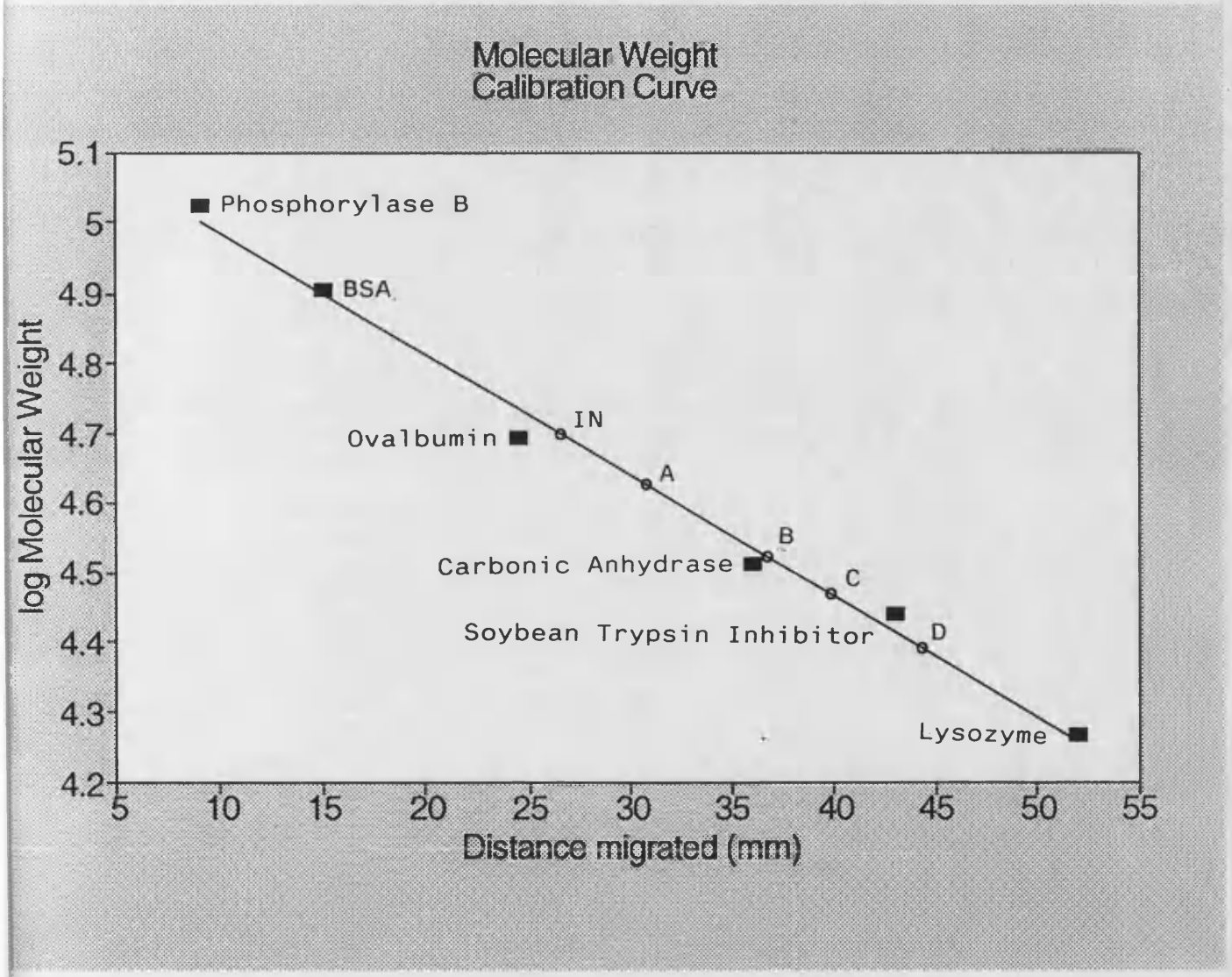

Molecular Weight
Calibration Curve
log Molecular Weight
5.1
5
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
Phosphorylase B
BSA
IN
Ovalbumin
A
B
Carbonic Anhydrase
C
Soybean Trypsin Inhibitor
D
Lysozyme
5
10
15
20
25
30
35
40
45
50
55
Distance migrated (mm)

Figure 4.3 Flow chart depicting the solubilization experiments used to denature and refold IN. NaCl, urea, an guanidine were buffered in 50 mM Tris-Cl, pH 8.0. Buffer A is 50 mM Tris-Cl, 1 mM DTT, 0.1 mM DTPA, 10% (w/v) glycerol pH 7.5. Buffer R is 50 mM Tris-Cl, 2 mM EDTA, 2 mM DTT, 0. M NaCl, 0.1% (v/v) Nonidet P-40, 10% (w/v) glycerol; pH 7.5).

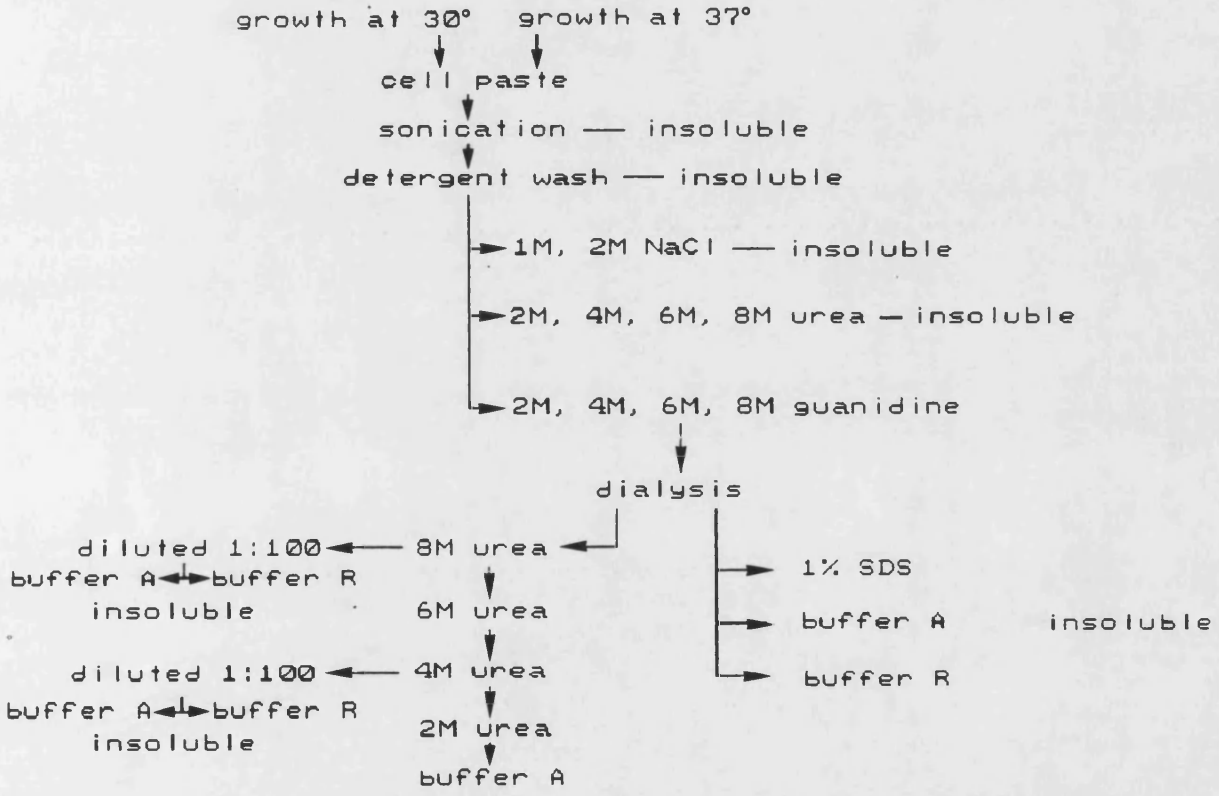

growth at 30°
growth at 37°
cell paste
sonication — insoluble
detergent wash — insoluble
1M, 2M NaCl — insoluble
2M, 4M, 6M, 8M urea — insoluble
2M, 4M, 6M, 8M guanidine
dialysis
8M urea
diluted 1:100
buffer A ⇄ buffer R
insoluble
6M urea
4M urea
diluted 1:100
buffer A ⇄ buffer R
insoluble
2M urea
buffer A
1% SDS
buffer A
buffer R
insoluble

Figure 4.4 Coomassie stained SDS-PAGE gel of sonication and wash. *E. coli* transformed with pIF was grown at 30°C, induced, sonicated and washed as described in section 2.2.21. Samples were resuspended in reducing SDS-PAGE sample buffer and loaded on a 7.5% acrylamide gel. Staining was with Coomassie Blue. The arrow indicates the position of IN.

Lanes 1-5 are TG1 transformed with pIF.

Lane 1 Whole cell lysate

Lane 2 Soluble fraction after wash

Lane 3 Insoluble fraction after wash

Lane 4 Soluble fraction after sonication

Lane 5 Whole cell lysate

Lane 6 Untransformed TG1 whole cell lysate

Lane 7 Molecular weight standards (kDa)

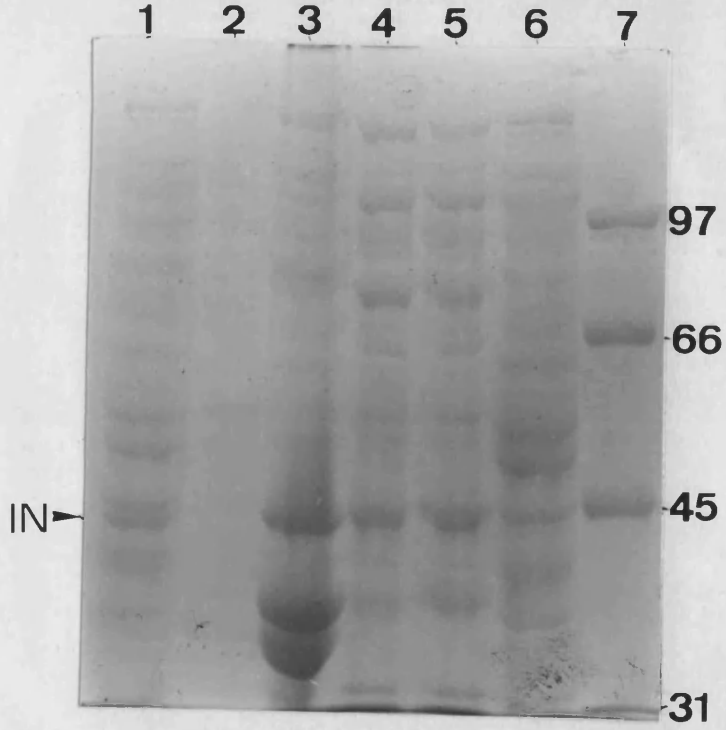

1
2
3
4
5
6
7
97
66
45
31
IN

Figure 4.5 Immunoblot of 6M and 8M urea solubilized IN. TG1 transformed with either pTZ19R or pIF were solubilized in urea as described in section 4.2.1. Samples were taken up in reducing SDS-PAGE sample buffer and separated on a 10% acrylamide gel. Immunoblotting was with a 1:400 dilution of anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualization was wit NBT/BCIP. The arrow indicates the position of IN.

Odd lanes are TG1 transformed with pTZ19R.
Even lanes are TG1 transformed with pIF.

Lane 1 pTZ19R 6M urea insoluble
Lane 2 pIF 6M urea insoluble
Lane 3 pTZ19R 6M urea soluble
Lane 4 pIF 6M urea soluble
Lane 5 pTZ19R 8M urea insoluble
Lane 6 pIF 8M urea insoluble
Lane 7 pTZ19R 8M urea soluble
Lane 8 pIF 8M urea soluble
Lane 9 Molecular weight standards (kDa)

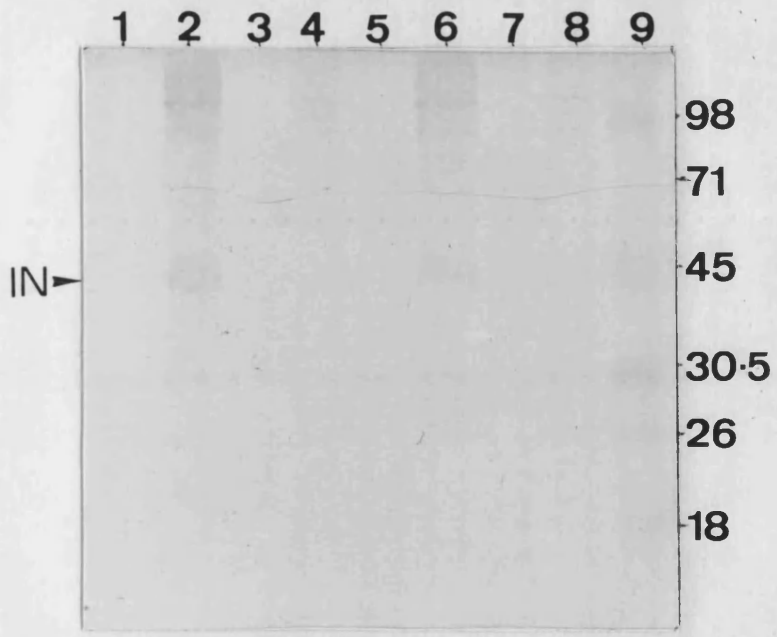
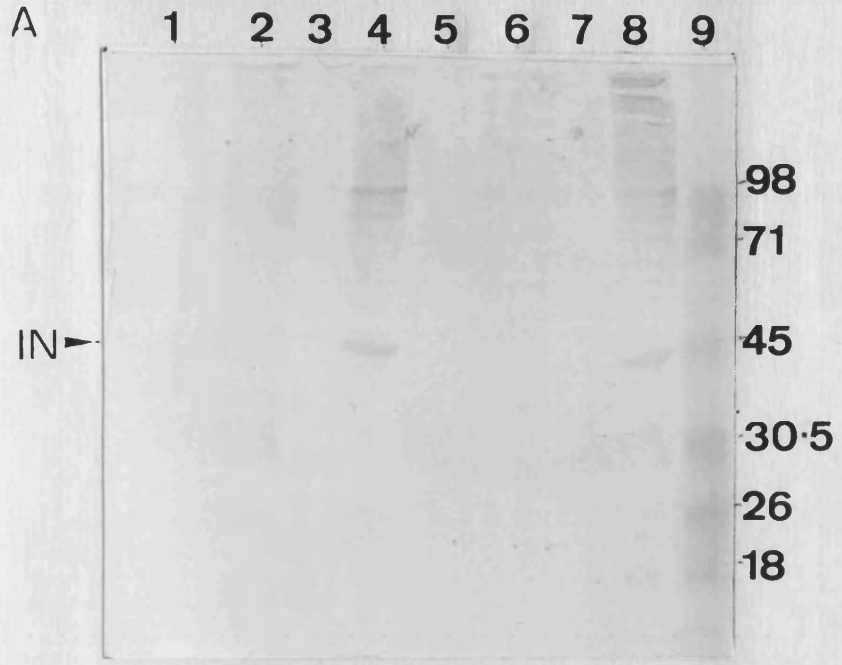

1 2 3 4 5 6 7 8 9
IN
98
71
45
30·5
26
18

Figure 4.6 Immunoblot of 2M, 4M, 6M, and 8M solubilization of IN in guanidine. TG1 transfomed with either pTZ19R or pIF were solubilized in guanidine as described in section 4.2.1. Samples were taken up in reducing SDS-PAGE sample buffer and separated on a 10% gel. Immunoblotting was with a 1:100 dilution of anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualization was with NBT/BCIP. The arrow indicates the position of IN.

In both A and B, odd lanes are TG1 transformed with pTZ19R and even lanes are TG1 transformed with pIF.
(guan=guanidine; insol=insoluble; sol=soluble)

A	B
Lane 1 pTZ19R 2M guan insol.	Lane 1 pTZ19R 6M guan insol
Lane 2 pIF 2M guan insol	Lane 2 pIF 6M guan insol
Lane 3 pTZ19R 2M guan sol	Lane 3 pTZ19R 6M guan sol
Lane 4 pIF 2M guan sol	Lane 4 pIF 6M guan sol
Lane 5 pTZ19R 4M guan insol	Lane 5 pTZ19R 8M guan insol
Lane 6 pIF 4M guan insol	Lane 6 pIF 8M guan insol
Lane 7 pTZ19R 4M guan sol	Lane 7 pTZ19R 8M guan sol
Lane 8 pIF 4M guan insol	Lane 8 pIF 8M guan sol
Lane 9 Mol. Wt. standards (kDa)	Lane 9 Mol. Wt. standards (kDa)


A
1 2 3 4 5 6 7 8 9
IN
98
71
45
30·5
26
18

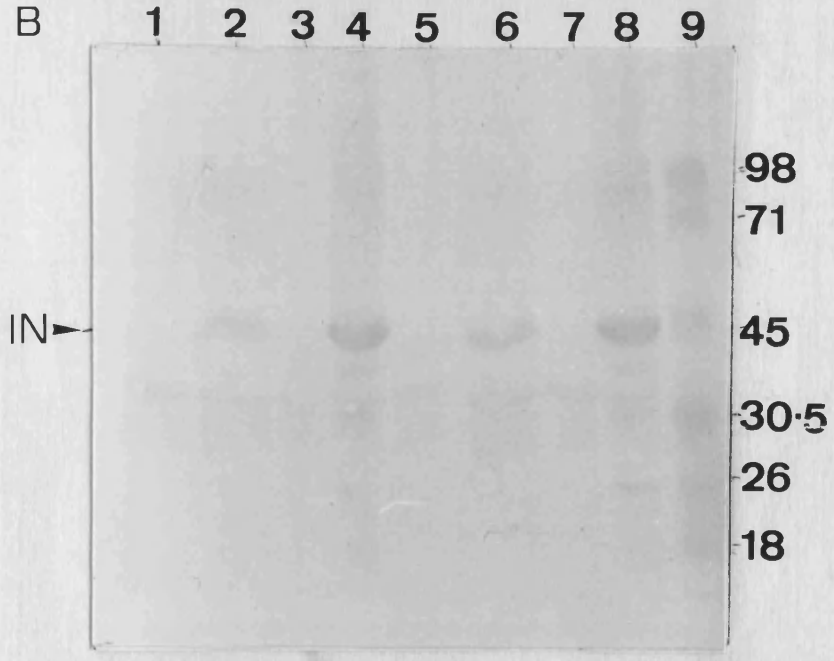
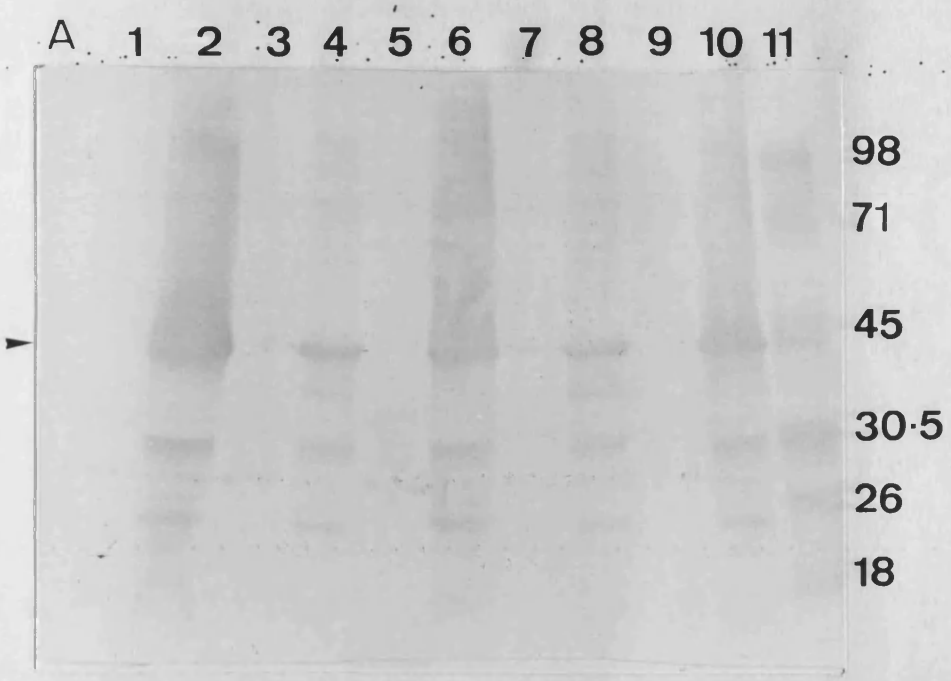

B
1 2 3 4 5 6 7 8 9
IN
98
71
45
30·5
26
18

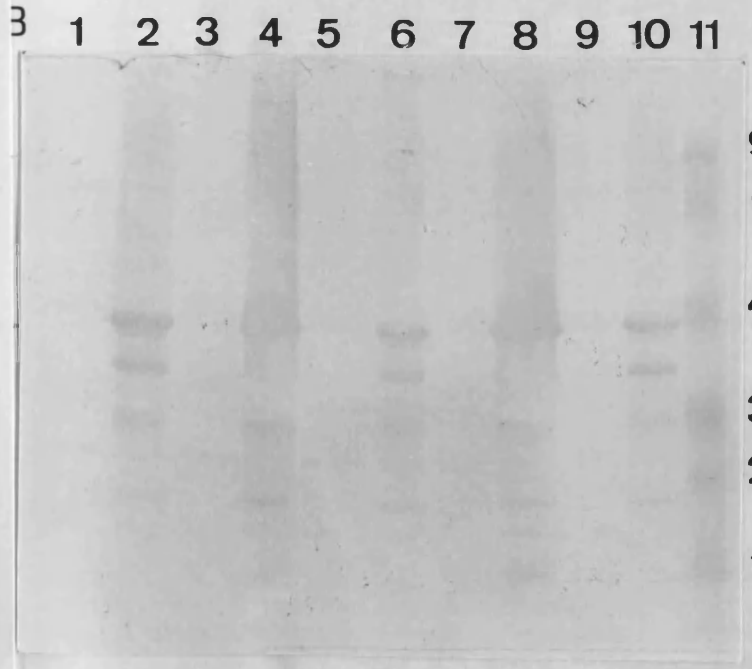
Figure 4.7 Immunoblot of 6M guanidine preparation of IN dialyzed against decreasing concentrations of urea. Sample soluble in guanidine were dialyzed sequentially against 8M, 6M, 4M, 2M urea, then buffer only. Samples were taken up i reducing SDS-PAGE sample buffer and separated on a 10% acrylamide gel. Immunoblotting was with a 1:100 dilution anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualization was wit NBT/BCIP. The arrow indicates the position of IN.

In both A, B, and C odd lanes are TG1 transformed with pTZ19R and even lanes are TG1 transformed with pIF. (insol=insoluble; sol=solulble)

A	B
Lane 1 pTZ19R 8M urea insol	Lane 1 pTZ19R 4M urea sol
Lane 2 pIF 8M urea insol	Lane 2 pIF 4M urea sol
Lane 3 pTZ19R 8M urea sol	Lane 3 pTZ19R 2M urea insol
Lane 4 pIF 8M urea sol	Lane 4 pIF 2M urea insol
Lane 5 pTZ19R 6M urea insol	Lane 5 pTZ19R 2M urea sol
Lane 6 pIF 6M urea insol	Lane 6 pIF 2M urea sol
Lane 7 pTZ19R 6M urea sol	Lane 7 pTZ19R buffer insol
Lane 8 pIF 6M urea sol	Lane 8 pIF buffer insol
Lane 9 pTZ19R 4M urea insol	Lane 9 pTZ19R buffer sol
Lane 19 pIF 4M urea insol	Lane 10 pIF buffer sol
Lane 11 Mol. Wt. standards	Lane 11 Mol Wt. standards (k

C. Coomassie-stained gel. As B lanes 7-11.


A
1 2 3 4 5 6 7 8 9 10 11
98
71
45
30·5
26
18


B
1 2 3 4 5 6 7 8 9 10 11
98
71
45
30·5
26
18

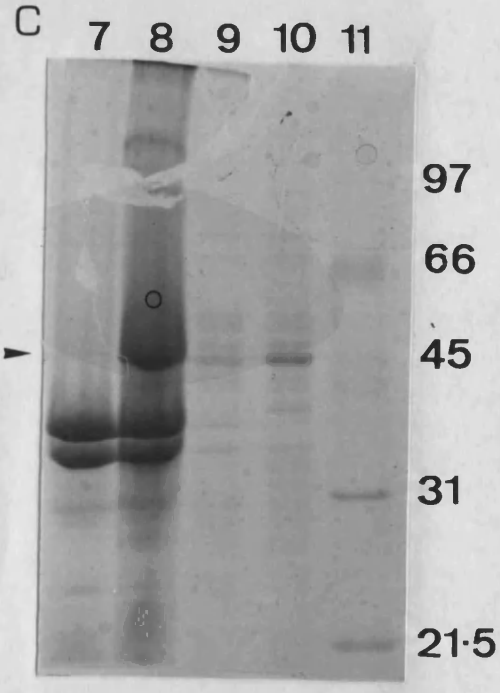

C
7 8 9 10 11
97
66
45
31
21·5

Figure 4.8 Immunoblot of expression of IN in *E. coli* BL21. BL21 transformed with pIF was induced after 2 or 3 hours of growth. Cells were harvested at 1, 2, or 4 hours after induction. Cells were taken up in reducing SDS-PAGE sample buffer and separated on a 10% acrylamide gel. Immunoblotting was with a 1:100 dilution anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualization was with NBT/BCIP. The arrow indicated the position of IN.

Lane 1 Whole cell lysate BL21 transformed with pTZ19R

Lane 2 (pIF) induction after 2 hours, 1 hour growth

Lane 3 Induction after 2 hours, 2 hour growth

Lane 4 Induction after 2 hours, 4 hour growth

Lane 5 Induction after 3 hours, 1 hour growth

Lane 6 Induction after 3 hours, 2 hour growth

Lane 7 Induction after 3 hours, 4 hour growth

Lane 8 Soluble fraction after sonication

Lane 9 Soluble fraction after wash

Lane 10 Molecular weight standards (kDa)

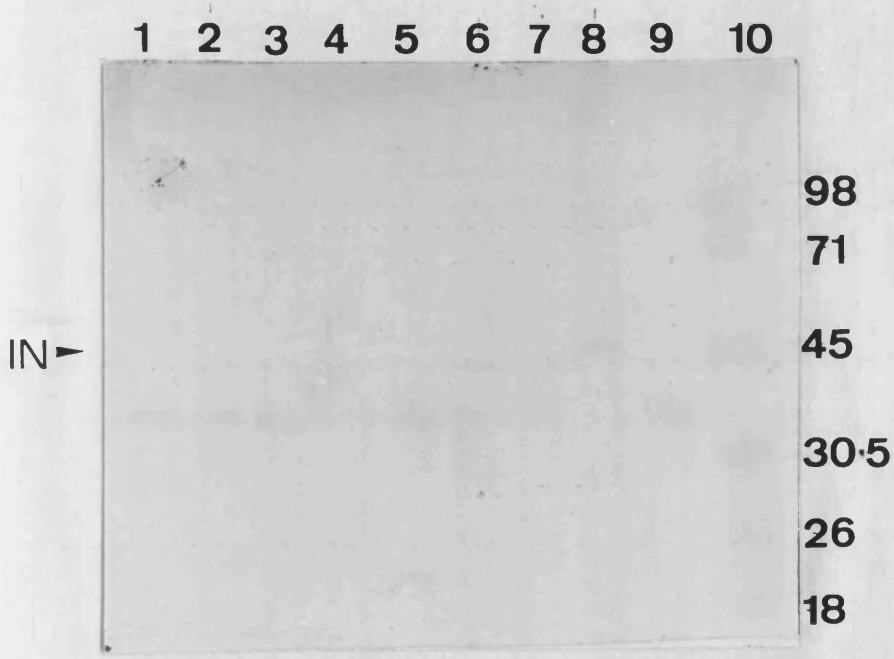
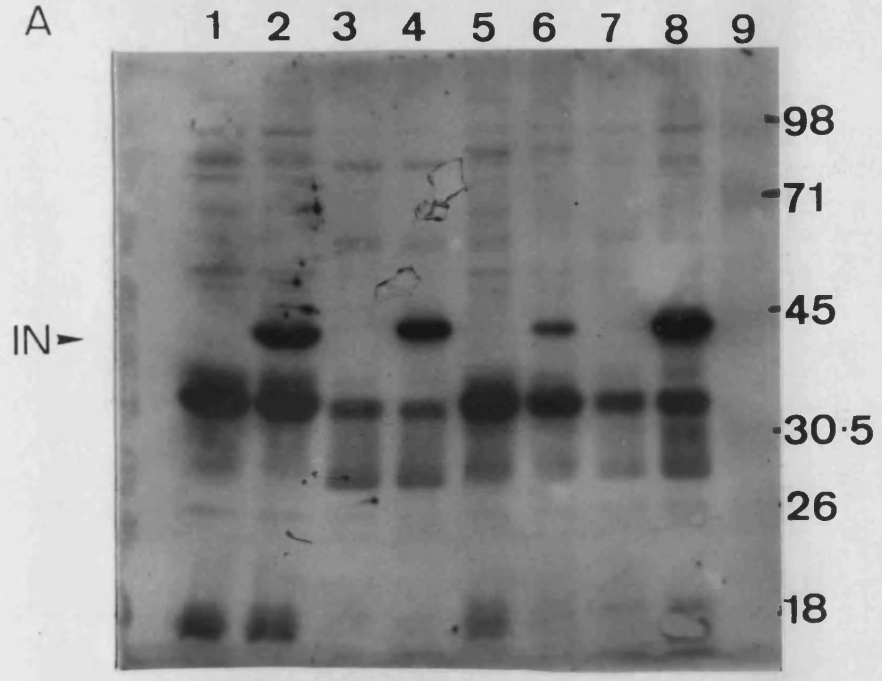
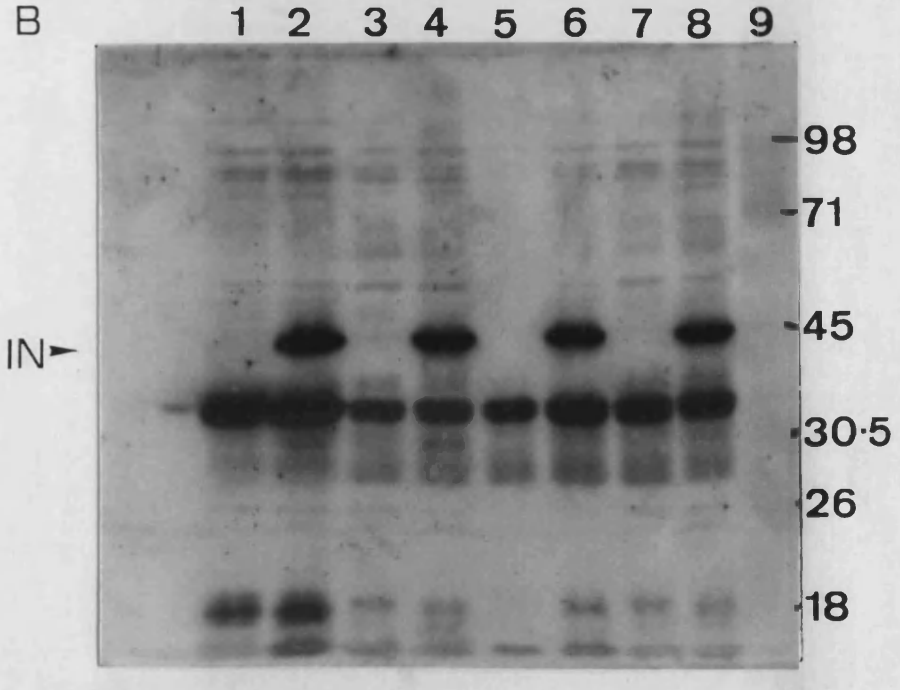

1 2 3 4 5 6 7 8 9 10
IN
98
71
45
30·5
26
18

Figure 4.9 Immunoblot of CAG629 transformed with either pTZ19R or pIF and solubilized in urea as described in section 4.2.1. Samples were taken up in reducing SDS-PAGE sample buffer and separated on a 10% acrylamide gel. Immunoblotting was with a 1:100 dilution of anti-KK28 anti-serum. The second antibody was a 1:1000 dilution of goat anti-rabbit IgG horseradish peroxidase conjugate. Visualization was with the ECL Detection system. The anti-KK28 anti-serum was not pre-absorbed.

In both A and B odd lanes are CAG629 transformed with pTZ19R and even lanes are CAG629 transformed with pIF.
(insol=insoluble; sol=soluble)

A	B
Lane 1 pTZ19R 2M urea insol	Lane 1 pTZ19R 6M urea insol
Lane 2 pIF 2M urea insol	Lane 2 pIF 6M urea insol
Lane 3 pTZ19R 2M urea sol	Lane 3 pTZ19R 6M urea sol
Lane 4 pIF 2M urea sol	Lane 4 pIF 6M urea sol
Lane 5 pTZ19R 4M urea insol	Lane 5 pTZ19R 8M urea insol
Lane 6 pIF 4M urea insol	Lane 6 pIF 8M urea insol
Lane 7 pTZ19R 4M urea sol	Lane 7 pTZ19R 8M urea sol
Lane 8 pIF 4M urea sol	Lane 8 pIF 8M urea sol
Lane 9 Mol. Wt. standards (kDa)	Lane 9 Mol. Wt. standards (kDa)


A
1 2 3 4 5 6 7 8 9
IN
98
71
45
30·5
26
18


B
1 2 3 4 5 6 7 8 9
IN
98
71
45
30·5
26
18

4.4 Discussion

An expression system for MoMLV IN was developed but the protein was expressed in insoluble inclusion bodies, a problem often encountered when recombinant proteins are expressed intracellularly in *E. coli* (Marston, 1986) and common for IN. Inclusion bodies first require separation from bacterial proteins, and this was partially accomplished in this study by sonication and detergent wash. Next, the inclusion bodies must be denatured. In the case of IN, a variety of denaturants have been employed, such as 1.0% SDS for HIV-1 and MoMLV (Hizi and Hughes, 1988), 1M NaCl for HIV-1 (Sherman and Fyfe, 1990), 8M urea for HIV-1 (Marcus-Sekura *et al.*, 1990) and 7M guanidine for MoMLV (Roth *et al.*, 1990). In the case of MoMLV IN expressed from plasmid pIF, 6M guanidine solubilized a proportion of the recombinant protein. Once the recombinant protein has been denatured, refolding is required and several conditions were investigated. IN consistently reaggregated and came out of solution in most of those conditions. The value of the anti-sera in detecting small quantities of IN was demonstrated with the observation of IN remaining in solution following step-wise dialysis against decreasing concentrations of urea.

As there can be variation in expression in different bacterial strains, BL21 and CAG629 were investigated. Expression in both was at very low levels, and although IN expressed in BL21 seemed to be soluble after sonication and

in CAG629 the protein did display solubility in urea, the experiments were not pursued. Expression in CAG629 also produced fewer degradation products; degradation will be discussed in Chapter 5.

Two factors are important in renaturation: correct disulphide bond formation and proline *cis-trans* isomerization. MoMLV has only three cysteines, two of which are implicated in co-ordinating zinc (see Chapter 7). Perhaps a more important consideration for MoMLV IN is proline content. *In vivo* prolines are all translated in the *trans* configuration, but some do assume the *cis* configuration in the folded protein. MoMLV IN has 32 prolines, a proline content of 7.8%. Comparison of soluble and insoluble proteins expressed as a fusion with cII protein in *E. coli* reveals that low proline content may correlate with solubility (Schein, 1989). Proteins with high proline content, however, show variable levels of solubility; for example, human c-myc (7.1% proline) is insoluble as well as human α-globin (5.0% proline) and *Xenopus* histones (5.6% proline). However, human plasma gelsolin (5.0% proline) and the C-terminus of human plasma gelsolin (5.4% proline) can be expressed as soluble proteins in *E. coli* (cited in Schein, 1989). In the case of these proteins, high proline content alone does not account for insolubility, with Ca^{2+} binding, cysteine content and repeats of negatively charged residues possibly playing a role.

Prolyl isomerases are ubiquitous and *E. coli* possesses two: PPIase A, active in the periplasm, and PPIase B, active in the cytoplasm. It would be expected that PPIase B should catalyze proline isomerization in cytoplasmically expressed recombinant proteins *in vivo.* However, given that i) isomerization is slow, ii) the rate of synthesis of recombinant protein is high, iii) polypeptides with incorrect proline bonds aggregate and iv) accessibility of the prolyl bond to the isomerase is important, aggregation possibly hindering access, all imply that only a small proportion of recombinant protein would possess the proper proline bonds. Provision of PPI during *in vitro* refolding or co-expressed *in vivo* are strategies that could be enployed to increase the amount of properly folded protein. That PPI *in vitro* can catalyze isomerization is well established, though not for all proteins (Lin *et al.*, 1988). Co-expression with PPI has not been studied extensively; one report has been found in the literature and that demonstrated no increase in correctly folded antibody fragments co-expressed with PPIase A (Knappik *et al.*, 1993). As with *in vitro* refolding, the success of co-expression may prove to be protein dependent.

4.5 Conclusions

Expression of MoMLV IN was investigated in three different bacterial strains. In *E. coli* TG1 IN was aggregated in insoluble inclusion bodies. Although denaturation was

achieved in 6M guanidine, refolding in various conditions caused IN to come out of solution. Both *E. coli* BL21 and CAG629 produced IN with greater solubility than in TG1. At the time the work described in this chapter was largely completed, there were no reports of successful expression of IN in *E. coli* in the literature. A different expression strategy was developed and is presented in Chapter 5. Had this work continued, various denaturation conditions would have been investigated, such as the effect of pH, ionic environment, and the ratio of denaturant to protein as well as the use of detergents as denaturant. Renaturation conditions to be investigated would be maintenance of reduction, pH and the ratio of renaturant to protein. Provision of PPI during *in vitro* refolding or co-expression of PPI could also be investigated.

Chapter 5

This chapter describes the construction of pMIN, a vector that expresses IN as a fusion with Maltose Binding Protein (MBP) and the purification of MBP-IN.

5.1 Introduction

A promising approach to the expression of IN from *E. coli* is as an MBP-IN fusion, a method successfully utilized to express HIV-IN (Kulkosky *et al.*, 1992; Vink *et al.*, 1993) and avian IN (Kitamura *et al.*, 1992). Expression levels were high with no need for solubilization and subsequent purification on an amylose affinity column yielded up to 95% pure MBP-IN. Although a strategically-placed Factor Xa cleavage site at the N-terminus of IN can yield native IN, in all three reports above MBP-IN was active in an *in vitro* auto-integration assay as a fusion protein.

Another promising development is the first report of active MoMLV IN expressed from *E. coli* in which the C-terminal 28 amino acids are deleted and replaced by a hexa-histidine tail (Jonsson *et al.*, 1993). Purification was on a nickel affinity column in 4M urea and refolding was accomplished by dialysis against reduced buffer in decreasing concentrations of urea. IN produced in this manner was active *in vitro* in 3' processing, strand transfer, and disintegration assays.

5.2 Materials and Methods

5.2.1 Materials

Taq polymerase was from Northumbria Biologicals or Perkin Elmer. dNTPs were from Pharmacia. The MBP fusion purification reagents were from New England BioLabs. Monopotassium glutamate was from Fluka.

5.2.2 Methods

5.2.2.1 Polymerase Chain Reaction (PCR)

PCR was performed in a Perkin Elmer Cetus DNA Thermal Cycler. The reaction contained 10 pmol template DNA, 20 pmol of each primer, 5 mM each dNTP, 2.5 U Taq DNA polymerase and PCR buffer (0.1 M Tris-Cl (pH 8.3), 0.5 M KCl, 15 mM $MgCl_2$, 0.01% (w/v) BSA) in a total volume of 100 µl. Thermal cycles consisted of: 94°C for 1 minute to allow DNA strand separation, 37°C for 1 minute for primer annealing, and 2 minutes at 72°C for DNA strand extension by Taq polymerase. Following 33 cycles the temperature was maintained at 72°C for 10 minutes to allow complete strand extension. Samples were stored at -20°C until analyzed by agarose gel electrophoresis.

5.2.2.2 Amylose Column Chromatography

The maltose binding protein-integrase fusion (MBP-IN) was purified by affinity chromatography on an amylose resin. Columns were prepared with pre-swollen resin in 2 sizes: 1 cm^3 and 10 cm^3. The column was washed with 8 column volumes of column buffer at 2-3 ml min^{-1}. After sonication, the soluble fraction of *E. coli* cells transformed with pMAL-c or pMIN (the integrase expression vector) was loaded onto the column at a flow rate of 1 ml min^{-1} and the column was washed with 8 volumes of column buffer. Elution was with 10 mM or 100 mM maltose in column buffer, and 1 ml or 3 ml fractions were collected. Two column buffers were used: Tris based (10 mM Tris-Cl (pH 7.4), 200 mM NaCl, 1 mM EDTA) or phosphate based (10 mM $NaPO_4$ (pH 7.2), 0.5 M NaCl, 1 mM Na azide, 10 mM β-ME, 1 mM EGTA). Following elution the column was washed with 1 volume H_2O, 3 volumes 0.1% (w/v) SDS, 1 volume H_2O and 3 volumes column buffer. The columns were used a maximum of 5 times before replacement and were stored at 4°C.

5.2.2.3 Cleavage with Factor Xa

The MBP-IN fusion was cleaved with Factor Xa in a ratio of 1% (w/w) Factor Xa to fusion protein in column buffer (listed above) at room temperature. Aliquots were taken at 2, 4, 8 and 24 hours and analyzed by SDS-PAGE for complete cleavage. For the large scale preparation of protein the

samples were incubated overnight (14 hours) at room temperature with gentle agitation.

5.2.2.4 Solubilization of IN After Cleavage

Following cleavage with Factor Xa, IN was solubilized in 4 M urea, 10% (w/v) glycerol and 0.1% (v/v) Nonidet P-40 in column buffer. The solution was gently agitated for 1 hour at room temperature. IN was renatured by step-wise dialysis against decreasing concentrations of urea in 10 mM Hepes (pH 7.4), 1 mM DTT, 20% (v/v) glycerol, 0.1% (v/v) Nonidet P-40, 400 mM monopotassium glutamate. Dialysis was over 5 days against 2 l of each concentration of urea (3 M, 2 M, or 1 M) for 24 hours with one change, and the final dialysis against 3.5 l with no urea over 48 hours with 2 changes.

5.3 Results

5.3.1 Construction of pMIN

Plasmid pMIN was designed to preserve the Factor Xa cleavage site such that following cleavage the IN amino acid sequence would be maintained. The IN coding region was provided by PCR amplification from plasmid pIF using oligo primers designed to introduce an *Eco*RV (blunt-ended)(all oligo sequences are listed in Table 2.1) restriction site at the 5' end and a *Pst*I site at the 3' end (figure 5.1). Plasmid pMAL-c was prepared by digestion with *Stu*I (blunt-ended) and

*Pst*I, sites for which are located in the polylinker of pMAL-c. Four ligations were unsuccessfully performed using a variety of DNA purification techniques, summarized in Table 5.1.

In the 2nd, 3rd and 4th ligation experiments recombinants were obtained, but none were of the expected size of 7.3 kb (IN insert 1.2 KB, vector 6.1 kb). Commonly occuring sizes were 2.0, 3.0, 4.2, and 6.6 kb. Restriction digest mapping indicated that these smaller plasmids did include the IN coding insert, pointing to a deletion in the pMAL-c vector. To rule out recombination due to the blunt-ended ligation, oligo Eag was designed to avoid a blunt-ended ligation using the *Eag*I restriction site in the polylinker while preserving the native IN reading frame and Factor Xa cleavage site. Oligo Eag was used as the 5' primer in the PCR reaction with oligo Pst1 (figure 5.2). The fragment was purified from agarose by freeze-squeeze, digested with *Eag*I and *Pst*I, again purifed from agarose by freeze-squeeze, then passaged through a NENsorb column. The pMAL-c vector was prepared from a CsCl maxi-prep, digested with *Eag*I and *Pst*I, dephosphorylated, purified from agarose by freeze-squeeze, then passaged through a NENsorb column. To address the problem of possible lethal expression, the ligation reaction was transformed into JM105, a non-expression strain of *E. coli*. Six recombinants resulted and were mini-prepped. Restriction digest analysis revealed one of these to be of the correct size, with a fragment of 1.2

kb. Plasmid pMIN was maxi-prepped by the CsCl method, then transformed into *E. coli* TB1.

5.3.2 Expression of MBP-IN

Whole cell lysates were prepared of TB1 transformed with plamsids pMAL-c or pMIN and were analyzed by SDS-PAGE and immunoblotting (figure 5.3). The Coomassie stained SDS-PAGE gel revealed no obvious unique band in the pMIN lysate. The immunoblot, probed with anti-maltose binding protein anti-serum, revealed two unique bands in the pMIN lysate. The immunoblot was repeated with induced cell lysates and probed with anti-KK28 anti-serum (figure 5.4). Eleven distinct bands were visualized in the molecular weight range of 93-53 kD. Molecular weights were calculated for proteins A-K (figure 5.5) and are as follows; A - 93.5, B - 87.8, C - 80.0, D - 71.3, E - 69.7, F - 68.1, G - 63.6, H - 60.8, I - 58.1, J - 55.5, K - 53.0 kD. Band B corresponds to the expected molecular weight of 88 kD for MBP-IN fusion and was produced in the largest quantity. The induced cells were also analyzed by immunoblotting with anti-KK334 anti-serum. No bands unique to the pMIN lysate were identified.

To address the possibility of lethal expression, saturated cultures of TB1 tranformed with pMAL-c and pMIN were prepared and used to inoculate fresh culture media. Bacterial growth was monitored by OD_{550} (figure 5.6). The growth curves indicate that expression of MBP-IN is not lethal to the cells. The slower growth rate of the induced

pMIN probably results from the transcription and translation of the fusion protein.

5.3.3 Trial Purification and Cleavage of MBP-IN

Trial purification and cleavage followed most of the manufacturer's recommendations. Trials were first done to determine the appropriate buffer for sonication in order to produce the greatest quantity of soluble fusion protein (described in detail in figure 5.7). The greatest amount of soluble proteins overall, including degradation products, was found with Basic Buffer. Purification was therefore undertaken using Basic Buffer as the sonication buffer.

Trials were conducted to ensure that MBP-IN would bind to the amylose resin and elute with maltose; this was confirmed. A 1 cm^3 amylose column was prepared and used to purify MBP-IN from a 25 ml culture of TB1 expressing either MBP-IN or MBP. The results are presented in figure 5.8. Although a proportion of the MBP-IN did not bind to the column, the majority eluted in the second fraction.

Timed cleavage trials were conducted on an aliquot of fraction 2 (figure 5.9). The disappearance of the MBP-IN over time indicates that cleavage is occurring. However, the cleaved IN was difficult to visualize on immunoblots. The possiblity existed that the IN was precipitating out of solution upon cleavage. In an attempt to purify even a small amount of IN, the cleavage reaction mixture was passed through the amylose column to remove the MBP. A precipitate

was observed collecting at the top of the column accompanied by a loss of flow rate. Although not confirmed by SDS-PAGE analysis, the conclusion was made that the IN had indeed precipitated.

An attempt was made to purify the MBP-IN from the degradation products. After amylose column chromatography, fractions 2 and 3 were pooled and separated by gel filtration on a Superdex column. Fractions were collected in 3 ml volumes and screened for proteins by SDS-PAGE and immunoblotting. No proteins were visualized by either method; gel filtration was thus abandoned.

5.3.4 Large Scale Purification and Cleavage of MBP-IN

Large scale purification was done with some modifications from the small scale. Reducing conditions were maintained during column chromatography and following cleavage of the fusion protein, denaturation and renaturation were done by the method of Jonsson and co-workers (1993). A 10 cm^3 amylose column was prepared and used to purify MBP-IN from a 100 ml culture of *E. coli*. Fractions were collected in 3 ml volumes and screened for the presence of MBP-IN and MBP by SDS-PAGE and immunoblotting (figures 5.10 and 5.11). MBP was identified in fractions 3, 4, and 5 and were pooled and stored at -20°C for use as a control. MBP-IN was identified in fractions 3, 4, and 5, were pooled, the protein concentration determined, then incubated with Factor Xa overnight. Aliquots were taken to determine solubility

levels and the remainder solubilized in urea to a final concentration of 4M. Following incubation for 1 hour aliquots were again taken to check solubility and the remainder dialyzed against decreasing concentrations of urea in a step-wise manner in the buffer described by Jonsson *et al.* (1993). Samples were then stored at -20°C. Aliquots from the various stages of the denaturation and renaturation were analyzed by SDS-PAGE and immunoblotting (figure 5.12). The results showed that the majority of IN was precipitated following cleavage but did return to solution when denatured in 4M urea (no pellet of insoluble material was observed when an aliquot of the 4M urea solution was centrifuged at 13,000 rpm for 10 minutes). Following dialysis (renaturation) the samples were centrifuged at 13,000 rpm and the soluble and insoluble fractions analyzed by SDS-PAGE. All the proteins were visible in the insoluble fraction on the SDS-PAGE gel and immunoblot, indicating all of the IN had precipitated.

5.3.5 Sequencing of pMIN

The IN coding region was completely sequenced from the coding strand of pMIN. Two base changes were noted from the published MoMLV IN sequence (Shinnick *et al.*, 1980). Nucleotide 4630 was C instead of A resulting in an amino acid substitution of Ser for Tyr. Nucleotide 5355 was A instead of G, an amino acid change from Ala to Asp. To determine the source of these point mutations, plasmid pIF

was sequenced across these regions (sense strand only). The sequence of pIF matched that of Shinnick, indicating that these mutations arose during PCR amplification of the IN coding region (figure 5.13).

Table 5.1 Listing of the purification of the insert and plasmid in each of the ligation experiments performed.

Expt.	Insert Purification	Plasmid Purification
1	PCR digestion Prepagene ligation	Plasmid from mini-maxi prep digestion Prepagene ligation
2	PCR Phenol extr. Ethanol ppt. CL6B column digestion ligation	Same as Expt. 1
3	PCR Gene Clean II digestion ligation	Plasmid from CsCL maxi prep digestion phenol extr. ethanol ppt. ligation
4	PCR Gene Clean II digestion phenol extr. ethanol ppt. ligation	Same as Expt. 3

Figure 5.1 PCR amplification of IN fragment from pIF for cloning into pMAL-c. PCR was performed as described in section 5.2.2.1 using primers Eco and Pst1. One-tenth of the reaction volume was analyzed on a 1% agarose gel and shows the expected product of 1.3 kb.

Lane 1 *Hind*III digested lambda size markers

Lane 2 Complete reaction mixture

Lane 3 Reaction minus template

Lane 4 Reaction minus Eco primer

Lane 5 Reaction minus Pst1 primer

Lane 6 Reaction minus Taq DNA polymerase

Lane 7 *Hind*III digested lambda size markers

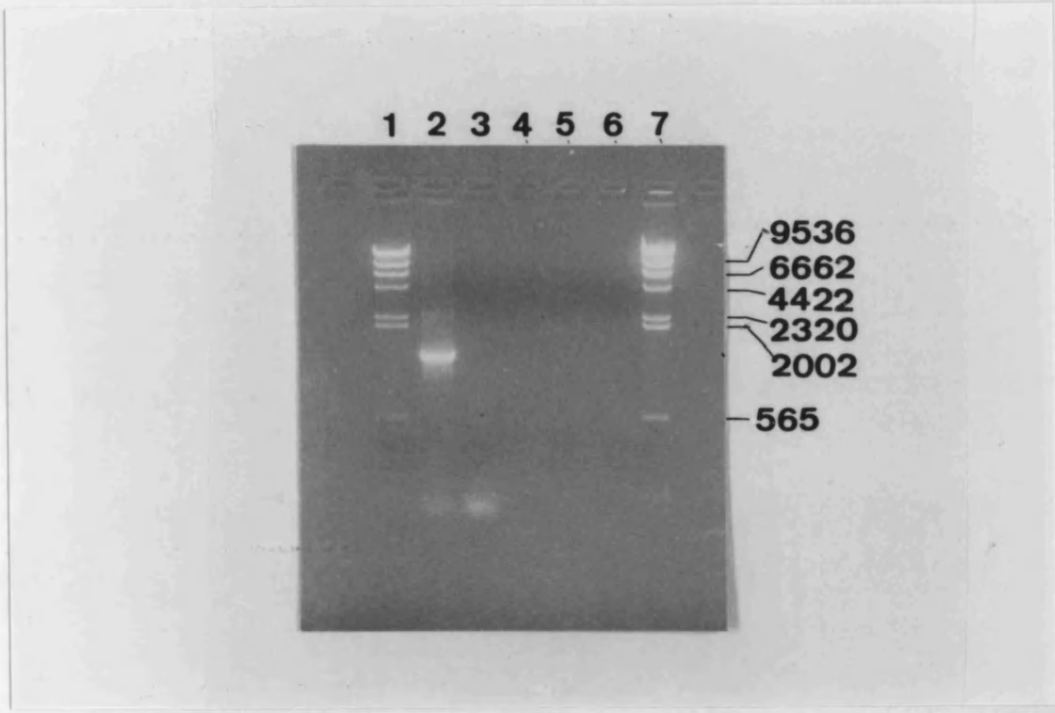

1 2 3 4 5 6 7
9536
6662
4422
2320
2002
565

Figure 5.2 PCR amplification of IN fragment from pIF for cloning into pMAL-c avoiding blunt-ended ligation. PCR was done as described in section 5.2.2.1 using primers Eag and Pst1. One-tenth of the reaction volume was analyzed on a 1% agarose gel and shows the expected product of 1.3 kb.

Lane 1 Complete reaction mixture

Lane 2 Reaction minus template

Lane 3 Reaction minus Eag primer

Lane 4 Reaction minus Pst1 primer

Lane 5 Reaction minus Taq DNA polymerase

Lane 6 *Pst*I digested lambda size markers

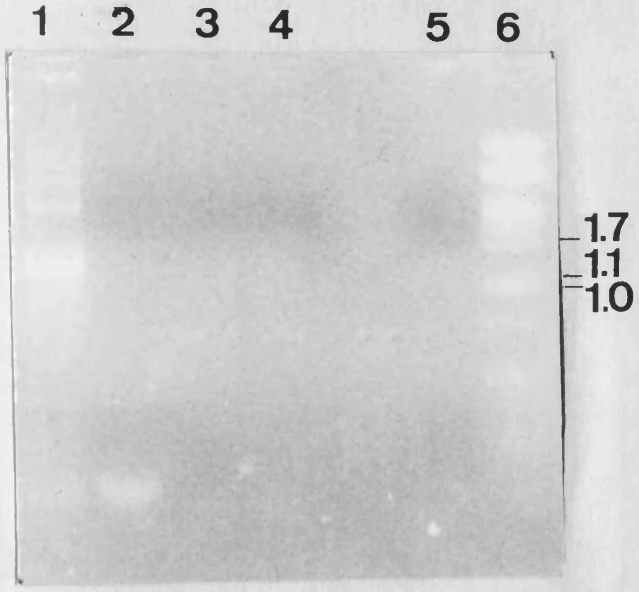

1
2
3
4
5
6
1.7
1.1
1.0

Figure 5.3 SDS-PAGE and immunoblot of uninduced cells expressing either pMAL-c or pMIN. A. Coomassie-stained ge of whole cell lysate preparations taken up in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. B. Immunoblot of the same whole cell lysate preparations as in A. The blot was probed with a 1:10,000 dilution of anti-MBP anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:5000 dilution. Visualization was with NBT/BCIP.

In A and B

Lane 1 TB1 transformed with pMAL-c

Lane 2 TB1 transformed with pMIN clone 1-2

Lane 3 TB1 transformed with pMIN clone 1-4

Lane 4 TB1 transformed with pMIN clone 5-4

Lane 5 TB1 transformed with pMIN clone 10-3

Molecular weight standards are in kDa.

The arrow indicates the position of MBP-IN.

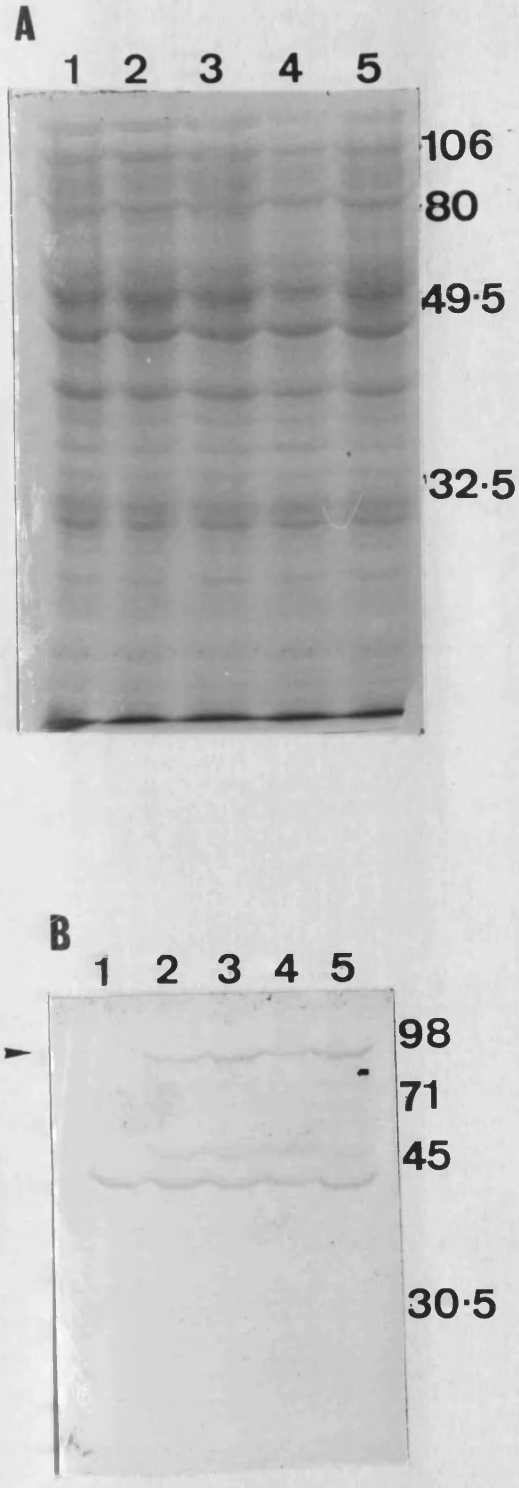

A
1 2 3 4 5
106
80
49·5
32·5
B
1 2 3 4 5
98
71
45
30·5

Figure 5.4 Immunoblot of induced cells expressing either pMAL-c or pMIN (clone 1-2). Whole cell lysate was resuspended in reducing SDS-PAGE sample buffer and separate on a 10% polyacrylamide gel. Immunoblotting was with a 1:400 dilution anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualizatio was with NBT/BCIP.

Lane 1 TB1 transformed with pMIN clone 1-2

Lane 2 TB1 transformed with pMAL-c

Lane 3 Molecular weight standards (kDa)

The arrow indicates the position of MBP-IN.

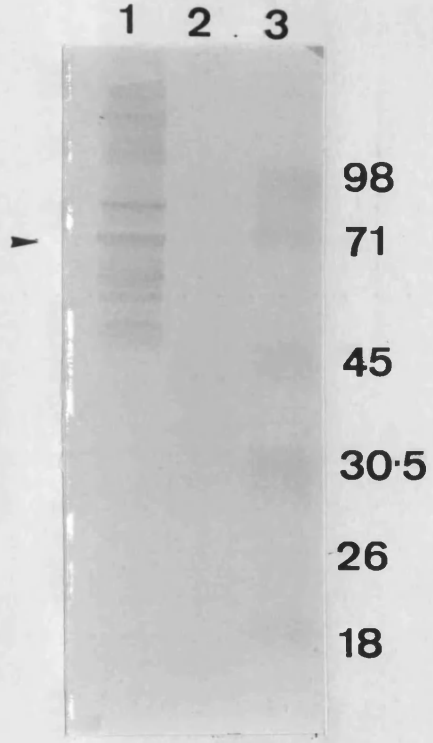

1
2
3
98
71
45
30·5
26
18

Figure 5.5 Calibration curve of the molecular weight of MBP-IN expressed in TB1, A-K are the major products identified from figure 5.4 lane 1.

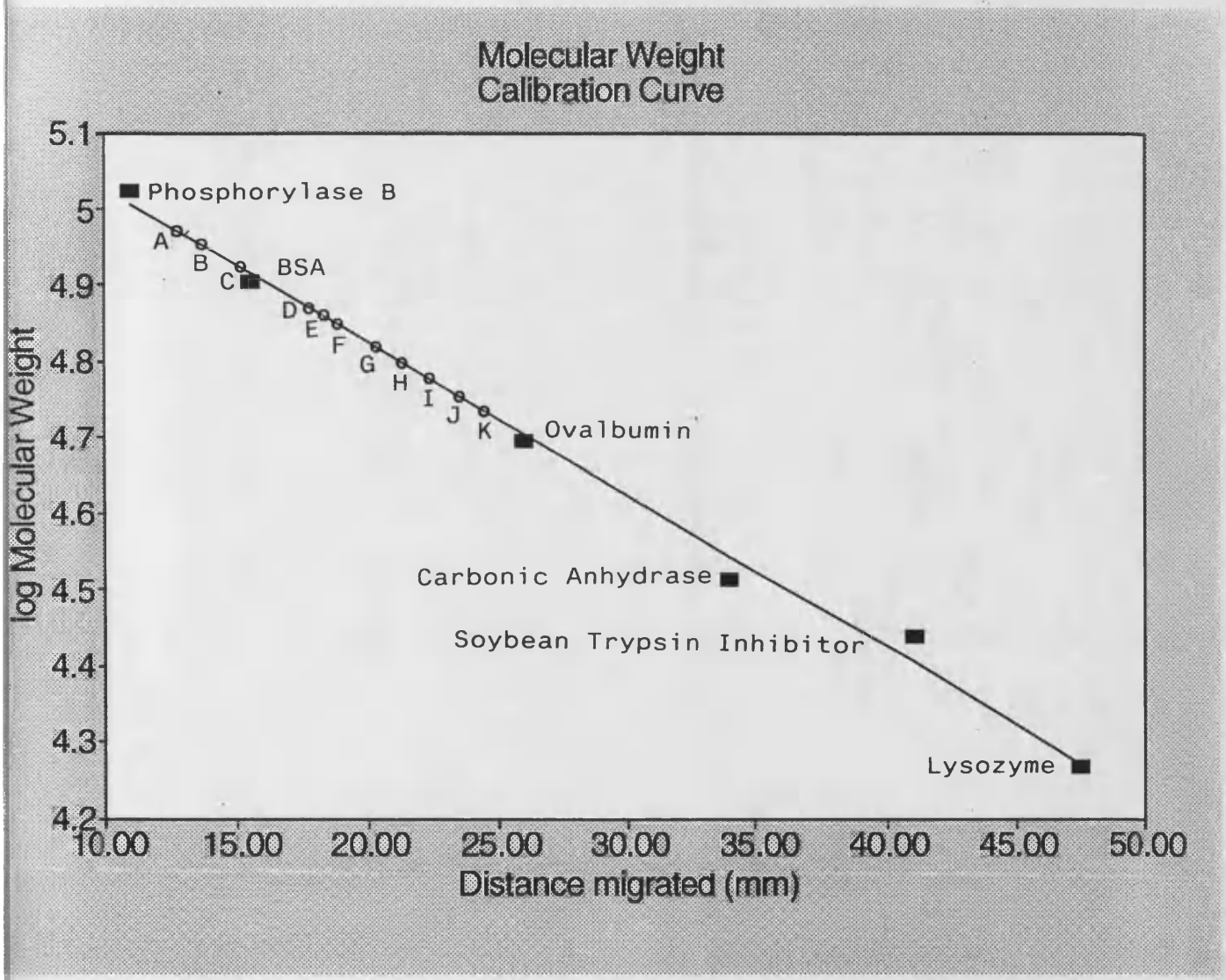

Molecular Weight
Calibration Curve
log Molecular Weight
5.1
5
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
Phosphorylase B
A
B
C
BSA
D
E
F
G
H
I
J
K
Ovalbumin
Carbonic Anhydrase
Soybean Trypsin Inhibitor
Lysozyme
10.00
15.00
20.00
25.00
30.00
35.00
40.00
45.00
50.00
Distance migrated (mm)

Figure 5.6 Growth curve of TB1 transformed with either pMAL c or pMIN and comparing pMIN induced and uninduced. Saturated cultures were used to inoculate fresh culture media. Growth was monitored by OD_{550}. Induction was with mM IPTG at 1.5 hours.

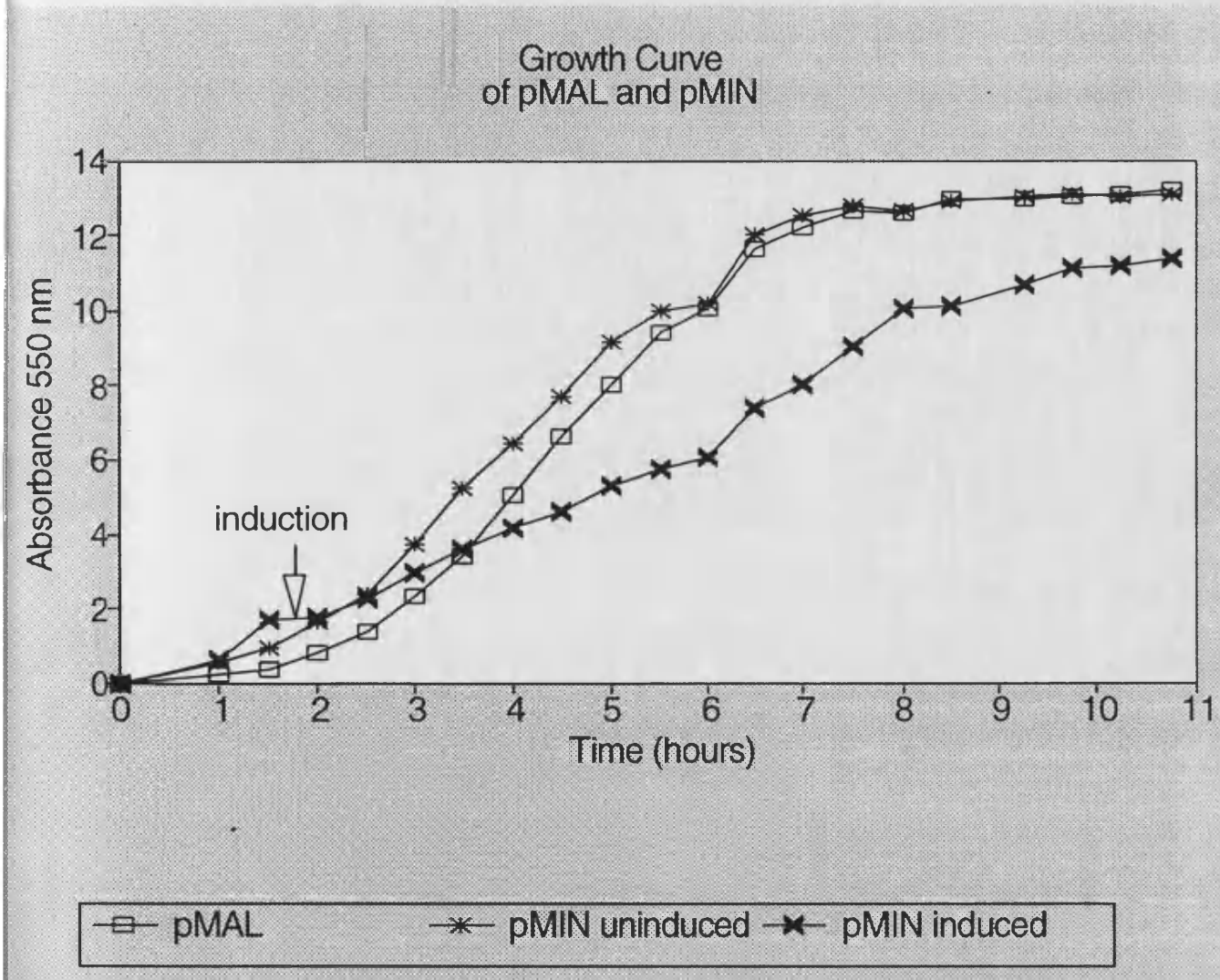

Growth Curve
of pMAL and pMIN
Absorbance 550 nm
14
12
10
8
6
4
2
0
induction
0
1
2
3
4
5
6
7
8
9
10
11
Time (hours)
pMAL
pMIN uninduced
pMIN induced

Figure 5.7 The effect of different sonication buffers on the quantity of soluble IN. TB1 transformed with pMAL-c or pMIN were grown and induced as described in section 2.2.22 and the cell debris pelleted by centrifugation at 13,000 rpm. Basic buffer is 50 mM Tris-Cl, 10 mM DTT, 1 mM DTPA; pH 7.5. Basic plus NaCl is Basic buffer plus 0.2 M NaCl. Lytic buffer is 10 mM $NaPO_4$, 30 mM NaCl, 0.25% (v/v) Tween-20, 10 mM β-mercaptoethanol, 10 mM EDTA, 10 mM EGTA; pH 7.0. Samples were resuspended in reducing SDS-PAGE sample buffer separated on a 10% polyacrylamide gel. Immunoblotting was with a 1:100 dilution anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was used at a 1:5000 dilution. Visualization was with NBT/BCIP.

Lane 1 Molecular weight standards (kDa)
Lane 2 Lytic pMIN soluble
Lane 3 Lytic pMAL soluble
Lane 4 Basic plus NaCl pMIN soluble
Lane 4 Basic plus NaCl pMAL soluble
Lane 5 Basic pMIN soluble
Lane 6 Basic pMAL soluble
Lane 7 Lytic pMIN insoluble
Lane 8 Lytic pMAL insoluble
Lane 9 Basic plus NaCl pMIN insoluble
Lane 10 Basic plus NaCl pMAL insoluble
Lane 11 Basic pMIN insoluble
Lane 12 Basic pMAL insoluble

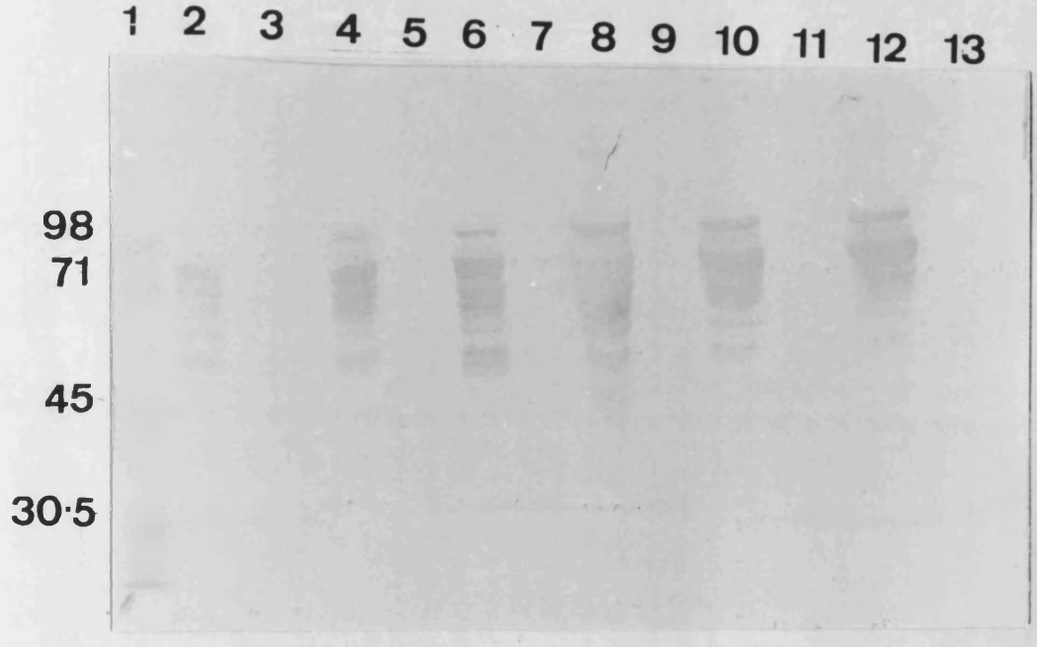

1 2 3 4 5 6 7 8 9 10 11 12 13
98
71
45
30·5

Figure 5.8 SDS-PAGE and immunoblot of 1 cm^3 amylose column purification. Samples of TB1 transformed with pMIN were sonicated and purified on the amylose column as described i section 5.2.2.2 with fractions collected as 1 ml volumes; elution was with 100 mM maltose. Aliquots of the fractions were resuspended in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. A. Coomassie-stained gel. B. Immunoblot with a 1:100 dilution anti-KK2 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was used at a 1:10,000 dilution. Visualization was with NBT/BCIP. The arrow indicates the position of MBP-IN.

In A and B

Lane 1 Void volume

Lane 2 Wash flow through

Lane 3 Elution fraction 1

Lane 4 Elution fraction 2

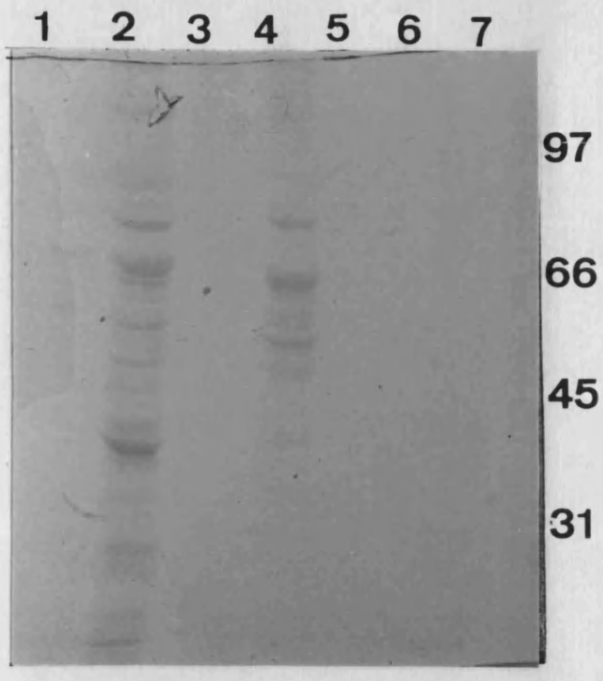
Lane 5 Elution fraction 3

Lane 6 Elution fraction 4

Lane 7 Elution fraction 5

Molecular weight standards are in kDa.

A



B

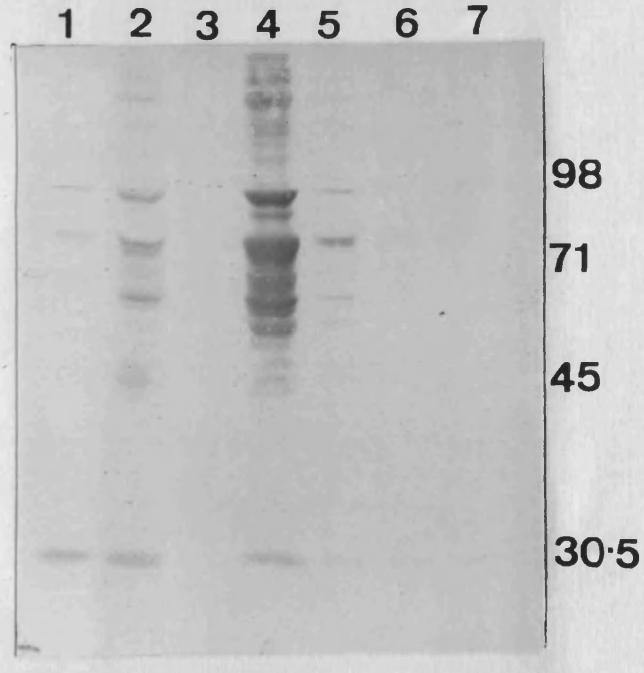


Figure 5.9 Immunoblot of Factor Xa cleavage trial. Cleavage was done as described in section 5.2.2.3 with an aliquot of elution fraction 2 (figure 5.8 lane 4). Aliquots were taken at 2, 4, 8, and 24 hours, resuspended in reducing SDS-PAGE sample buffer and separated on a 10% gel. Immunoblotting was with a 1:100 dilution of anti-KK28 anti-serum and a 1:10,000 dilution of goat anti-rabbit IgG alkaline phosphatase. Visualization was with NBT/BCIP.

Lane 1 Elution fraction 2

Lane 2 2 hour cleavage

Lane 3 4 hour cleavage

Lane 4 8 hour cleavage

Lane 5 24 hour cleavage

Molecular weight standards are in kDa.

The open arrow indicates the position of MBP-IN, the closed arrow indicates the position of IN.

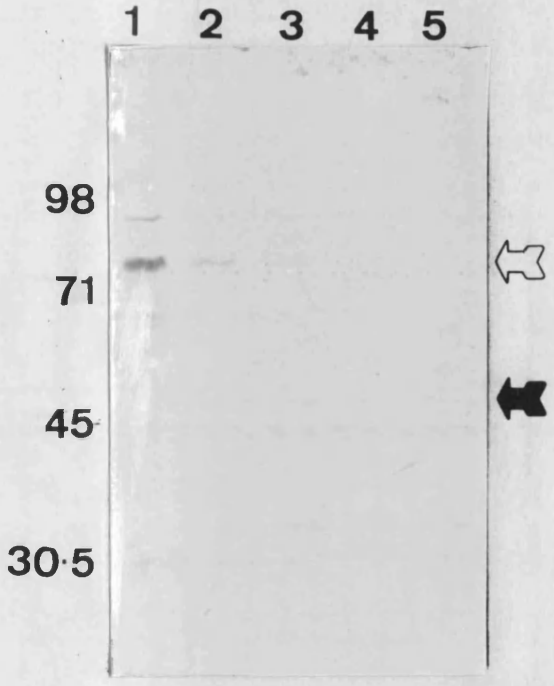

1
2
3
4
5
98
71
45
30·5

Figure 5.10 SDS-PAGE and immunoblot of MBP purification. TB1 transformed with pMAL-c was grown, sonicated and purified on the 10 cm^3 amylose column. Elution was with 10 mM maltose and 3 ml fractions were collected. Aliquots wer resuspended in reducing SDS-PAGE sample buffer and separate on a 10% polyacrylamide gel. A. Coomassie-stained gel. B Immunoblot with a 1:10,000 dilution anti-MBP anti-serum and a 1:10,000 dilution goat anti-rabbit IgG alkalkine phosphatase. Visualization was with NBT/BCIP.

In A and B

Lane 1 Void volume

Lane 2 wash flow through

Lane 3 Elution fraction 1

Lane 4 Elution fraction 2

Lane 5 Elution fraction 3

Lane 6 Elution fraction 4

Lane 7 Elution fraction 5

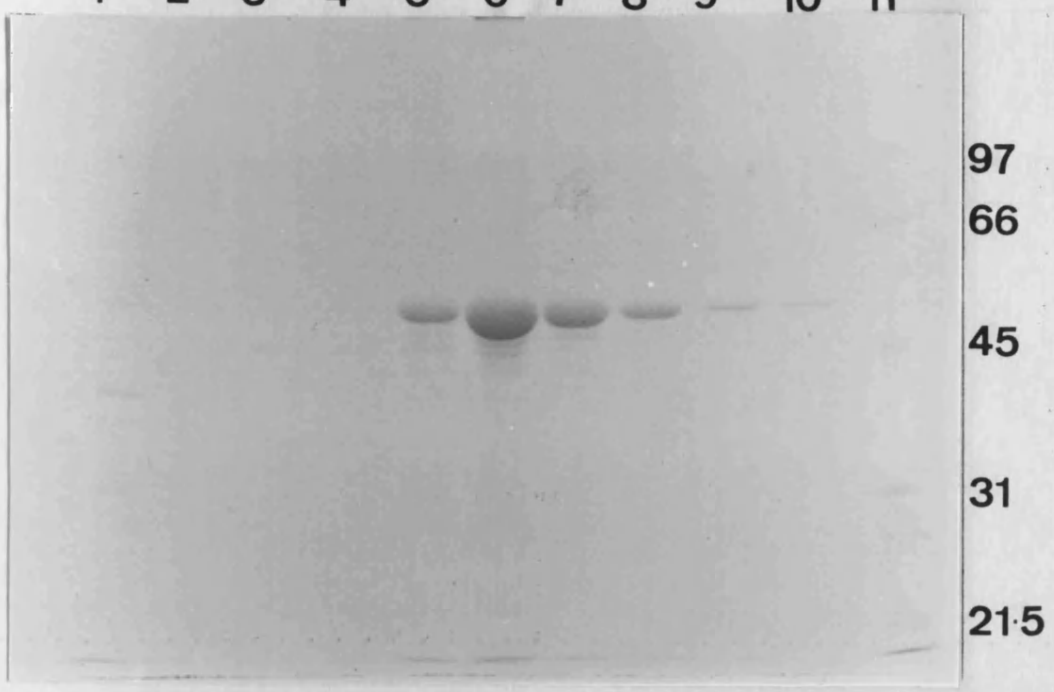
Lane 8 Elution fraction 6

Lane 9 Elution fraction 7

Lane 10 Elution fraction 8

In A Lane 11 Molecular weight standards (kDa)

A



B

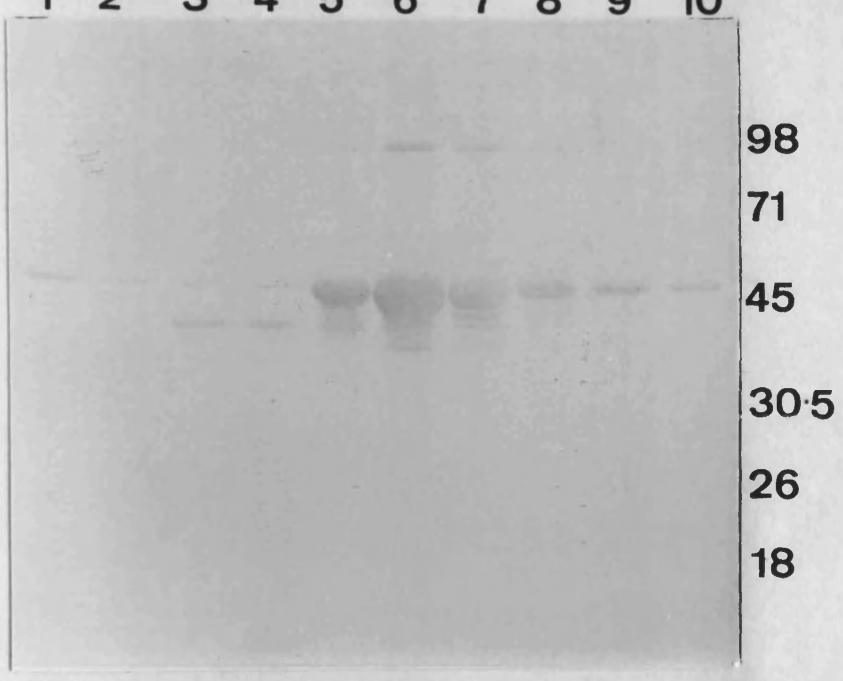


Figure 5.11 SDS-PAGE and immunoblot of MBP-IN purification TB1 transformed with pMIN was grown, sonicated and purified on the 10 cm^3 amylose column. Elution was with 10 mM maltose and 3 ml fractions were collected. Aliquots were resuspended in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. A. Coomassie-stained gel. B Immunoblot with a 1:100 dilution anti-KK28 anti-serum and a 1:10,000 dilution goat anti-rabbit IgG alkalkine phosphatase. Visualization was with NBT/BCIP.

In A and B

Lane 1 Void volume

Lane 2 wash flow through

Lane 3 Elution fraction 1

Lane 4 Elution fraction 2

Lane 5 Elution fraction 3

Lane 6 Elution fraction 4

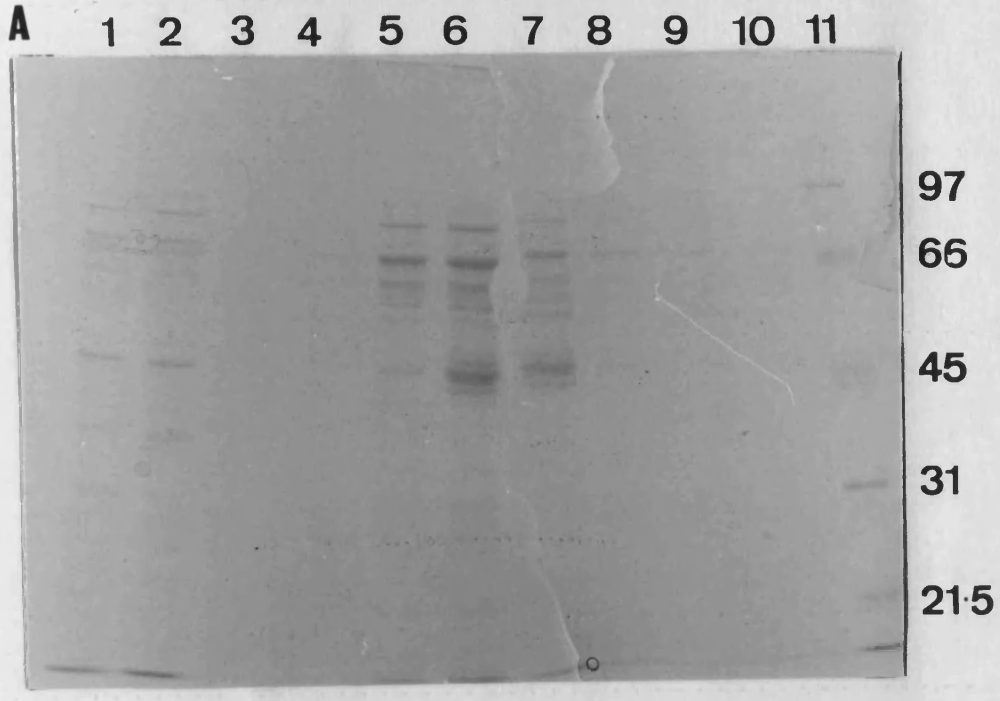
Lane 7 Elution fraction 5

Lane 8 Elution fraction 6

Lane 9 Elution fraction 7

Lane 10 Elution fraction 8

Lane 11 Molecular weight standards (kDa)


A
1 2 3 4 5 6 7 8 9 10 11
97
66
45
31
21·5

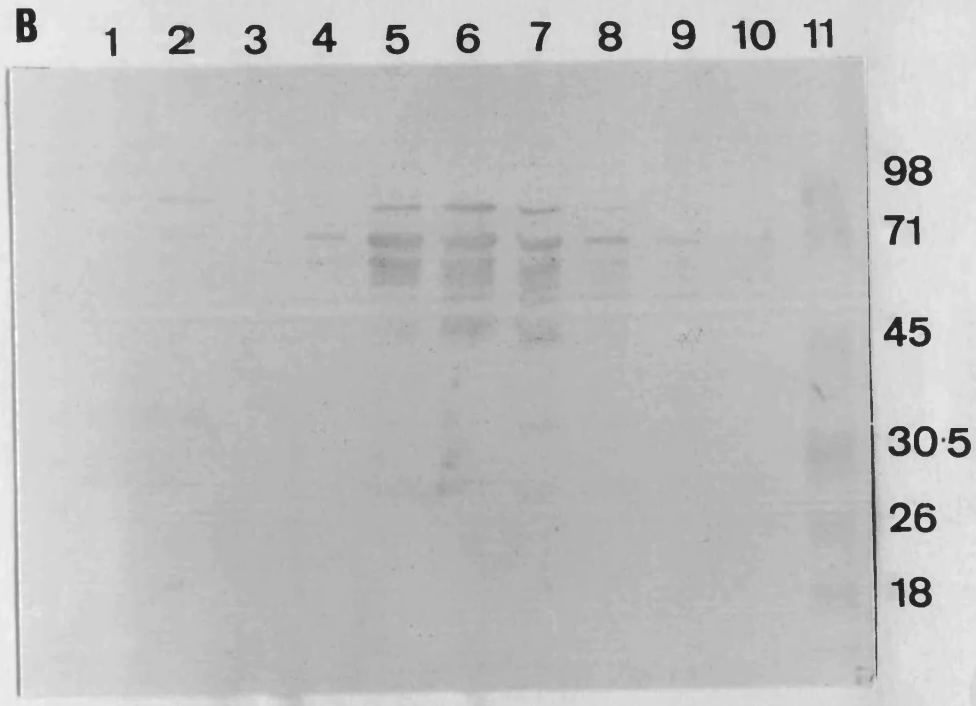

B
1 2 3 4 5 6 7 8 9 10 11
98
71
45
30·5
26
18

Figure 5.12 Immunoblot of MBP-IN cleavage with Factor Xa, solubilization in 4M urea, and following dialysis as described in section 5.2.2.4. Samples were resuspended in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. Immunoblotting was with a 1:100 dilution anti-KK28 anti-serum. Goat anti-rabbit IgG alkaline phosphatase was a 1:10,000 dilution. Visualizatio was with NBT/BCIP.

Lane 1 Elution fraction 4
Lane 2 Soluble fraction after cleavage
Lane 3 Soluble fraction in 4M urea
Lane 4 Insoluble fraction after dialysis
Lane 5 Soluble fraction after dialysis

Molecular weight standards are in kDa.

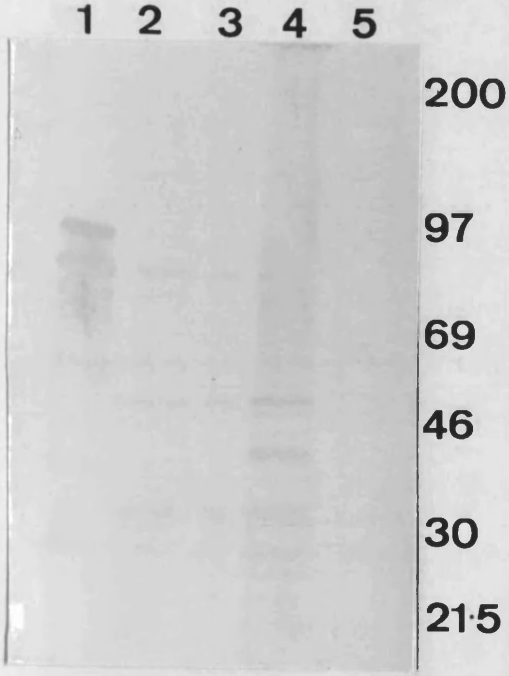

1 2 3 4 5
200
97
69
46
30
21·5

Figure 5.13 DNA sequence illustrating the mutations during PCR amplification of the IN coding region. Sequencing was done by the dideoxy chain-termination method (section 2.2.18).

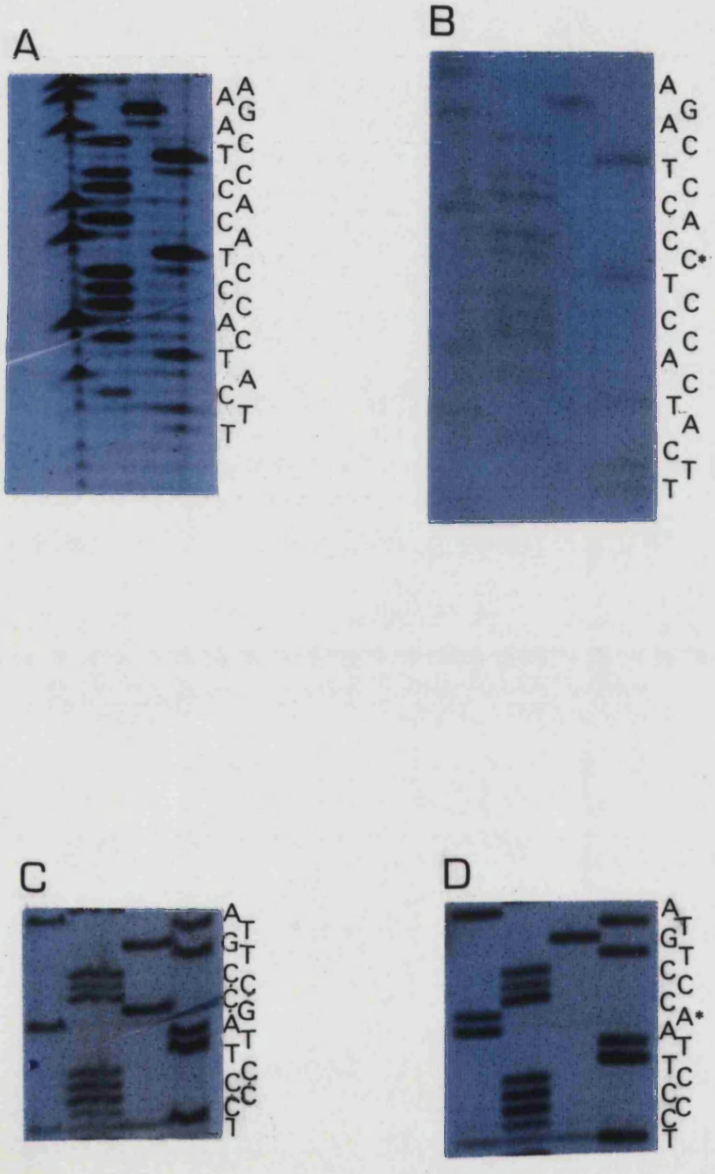
A. Residues 4619-4640 from pIF using Bam2 as primer

B. Residues 4619-4640 from pMIN using MalE as primer

C. Residues 5347-5362 from pIF using P5250 as primer

D. Residues 5347-5362 from pMIN using P5250 as primer

Starred residues indicate changes from the wild-type.


A
B
C
D

5.4 Discussion

5.4.1 Expression and Purification of MBP-IN

Vector pMIN was successfully constructed by avoiding the blunt-ended ligation and transformation into a non-expression strain of *E. coli*. Why the blunt-ended ligation consistently produced truncated vectors is unknown; lethal expression was ruled out, but recombination remains a possibility. One late discovery can explain some of the difficulty. Oligo Pst1, designed to create a *Pst*I site, was synthesized with two incorrect bases, resulting in the loss of the *Pst*I site. Although this explains why the construction of pMIN proved difficult, it does not explain the truncation of the vector. Sequencing across this region revealed the sequence expected if the *Pst*I site were intact. Sequencing the remainder of pMIN revealed two nucleotide mutations that occured during PCR amplification. Taq DNA polymerase has an error rate of 2×10^{-4} mutations per base duplication (Tindall and Kunkel, 1988; Keohavong and Thilly, 1989). At nucleotide 4630 a purine to pyrimidine substitution and at 5355 a purine to purine substitution. The resulting amino acid change at residue 7 of IN, Tyr to Ser, a non-conservative mutation may affect IN function. A four amino acid insertion between residues 11 and 12 of MoMLV IN was unable to form an integrated provirus *in vivo* (Donehower, 1988) while deletions of the N-terminus of HIV-1 IN (-16 or -38 residues) were inactive in 3' processing

and strand transfer assays *in vitro* (Schauer and Billich, 1992). The amino acid change at residue 249, Ala to Asn, also a non-conservative mutation, may also affect IN function. Although this mutation is not located in the central D, D(35)E region, important for all IN activites *in vitro*, it is only 29 amino acid residues from this region. One characterized mutant in this same region, a three amino acid insertion at position 239, was unable to form an integrated provirus *in vivo* (Donehower, 1988). Unfortunately, the late discovery of these mutations did not allow sufficient time for the vector to be constructed again.

MBP-IN was expressed to high levels and purified to near homogeneity. Following cleavage with Factor Xa, the evidence suggested that IN precipitated out of solution. This also suggests that IN was not properly folded despite being soluble as a fusion protein. Based upon this observation, together with the refolding considerations listed in Chapter 4 and the successful renaturation of MoMLV IN reported by Jonsson *et al* (1993), a large scale purification was undertaken with denaturation and refolding in the conditions outlined by Jonsson and co-workers. Unfortunately, this approach also met with failure, again suggesting that IN was not properly folded as part of the fusion protein, nor able to refold properly following cleavage. The reasons for IN's failure to fold are discussed in detail in Chapter 4.

5.4.2 Degradation Products

Both expression systems described produced degradation products. Four major products were identified from the pIF system and ten products from pMIN. Degradation products are the results of three processes: i) internal initiation, ii) premature termination of transcription or translation, and iii) proteolytic degradation. Others have reported degradation products in the *E. coli* expression of IN: HIV-1 (Vincent *et al.*, 1993), HIV-1 expressed as an MBP-IN fusion (Vink *et al.*, 1993), internal initiation products for MoMLV (Hizi and Hughes, 1988; Krogstad and Champoux, 1990) HIV-1 (Hizi and Hughes, 1988; Zervos *et al.*, 1990) and RSV (Mumm *et al.*, 1992). The anti-sera described in Chapter 3 aided in the identification of the degradation products produced from pIF. A 42 kD product is recognized by both anti-KK28 and anti-KK334 (see figure 3.7 A and B) and corresponds to the internal initiation product reported by Krogstad and Champoux (1990). The other degradation products identified in figure 3.7 are only recognized by anti-KK28 indicating that the C-terminal portion of the protein is missing. Unfortunately similar analysis for the pMIN products was unsuccessful as anti-KK334 did not recognize MBP-IN; this may be another indication that MBP-IN was not properly folded.

Recombinant protein expression can vary in different bacterial strains (Schoemaker *et al.*, 1985). Two different strains were investigated in this study: BL21, in addition

to producing soluble, active HIV-1 IN, is a *lon*⁻ strain, deficient in the La protease (Goff and Goldberg, 1985) and CAG629, a *htpR*⁻ strain. The gene product of this locus is involved in regulation of the heat shock response and also defective in proteolysis (Baker *et al.*, 1984). Although both expressed IN at low levels, CAG629 and BL21 seemed to produce fewer degradation products. Degradation was also limited in uninduced cells from pMIN (compare figure 5.3 and 5.4). It would have been valuable to compare by immunoblotting induced and uninduced cells for the pIF expression system to determine if degradation was reduced. Although degradation was reduced in protease-deficient strains and in the pMIN uninduced cell, the question remains if this was truly decreased degradation or lack of visualization of degradation products due to the lower expression levels. Another unresolved question is if any of the smaller products are due to premature termination of transcription or translation caused by secondary structure formation or rare codons. In comparing the degradation products of both expression systems, only two are in the same molecular weight range (26.0 from pMIN and 25.7 from pIF; 29.0 from pMIN and 29.7 from pIF). Also, an analysis of mRNA secondary structure reveals no obvious structures that could account for proteins of these sizes and since, the level of degradation seems to be strain dependent, it seems likely that these smaller products are due to proteolytic degradation.

Chapter 6

This chapter describes the efforts made to demonstrate recombinant IN activity and the development of a kinetic assay to measure 3' processing.

6.1 Introduction

Several methods can be employed to demonstrate IN activities *in vitro*. DNA binding can be demonstrated by gel mobility shifting (Basu and Varmus, 1990) or Southwestern blotting (Roth *et al.*, 1988). Endonuclease activity can be measured in a supercoiled plasmid nicking assay (Sherman and Fyfe, 1990; Marczinovitz *et al.*, 1992). Integration assays are either genetic or physical. The genetic assay is based upon genetic selection in *E. coli* using bacteriophage lambda gtWES (Brown *et al.*, 1987). The physical assay is based on radio-labelled duplex oligonucleotides that match the LTR ends. Following incubation with IN the oligo is analyzed on an acrylamide gel which detects 3' processing and the products of strand transfer as the oligo substrate can also serve as the target DNA (Craigie *et al.*, 1990).

Few kinetic assays have been developed. Although the oligonucleotide autointegration assay can be quantified by densitometric scanning, the kinetics are complicated as it is measuring both 3' processing and strand transfer. One assay, developed by Billich *et al.* (1992), measured the kinetics of 3' processing by radio-labelling the

dinucleotide cleaved in 3' processing and following incubation with HIV-1 IN, bound the oligo substrate to DEAE cellulose, washed the radiolabelled dinucleotide through, then measured the decrease in radioactivity. Velocity was linearly proportional to enzyme concentration. Integration activity of HIV-1 has been quantified by measuring a radio-labelled LTR substrate integrated into plasmid DNA (Carteau *et al.*, 1993). By isolating the plasmid DNA, separating on agarose gels, and densitometric analysis, velocity was shown to be linearly proportional to enzyme concentration, but also showed substrate inhibition.

The kinetic assay developed for this project is based in part on an assay described by Craigie and co-workers (1991) in which biotin-coupled, radio-labelled oligonucleotides are reacted with IN, then bound to avidin-coated microtitre plates. Autoradiography of the plates detects products of strand transfer. In collaboration with Amersham International, a continuous direct kinetic assay to measure 3'; processing was developed: the Scintillation Proximity Assay (SPA). The assay makes use of micron-sized beads impregnated with liquid scintillant molecules and coated with streptavidin. Biotinylated oligonucleotides with the two terminal thymidines tritiated (i.e. the dinucleotide to be cleaved) are reacted with IN and the decrease in counts measured (figure 6.1). Theoretically, the assay could also be developed to measure the kinetics of strand transfer and disintegration.

Figure 6.1 Schematic illustration of SPA. The bead is impregnated with liquid scintillant molecules and coated with streptavidin. The oligonucleotide is biotinylated and the two terminal nucleotides are tritiated (marked by filled circles). The principle of SPA relies on the fact that the energy from weak β-emitters (such as ^{3}H) will be dissipated into the aqueous medium unless the radioactive molecule is bound directly or indirectly to the bead. This can then stimulate the scintillant to emit light. Otherwise, no signal is generated. When reacted with IN, IN will cleave the thymidines, and a decrease in radioactivity will be measured.

6.2 Materials and Methods

6.2.1 Materials

Klenow polymerase and dNTPs were from Pharmacia. [α^{32}P]ATP was from NEN Dupont. Exonuclease III was from Boerhinger Mannheim. Biotinylated, tritiated oligos and streptavidin-coated scintillation beads were a gift from Amersham, Cardiff, U.K.

6.2.2 Methods

6.2.2.1 Southwestern Blotting

Southwestern Blotting was carried out by the method of Roth *et al.* (1988). Following protein separation by SDS-PAGE and transfer to nitrocellulose, the filter was blocked in either Blocking buffer [0.02% (w/v) BSA (boiled 10 minutes), 30 mM Hepes] or Blotto [5.0% (w/v) Marvel non-fat dry milk, 0.02% (w/v) sodium azide in PBST] for 1 hour at room temperature. The filter was washed in Buffer A (50 mM Tris-Cl, pH 8.3, 50 mM DTT, 1 mM EDTA, 8 M guanidine HCl) for 1 hour at room temperature, followed by washing with dilution buffer (50 mM Tris-Cl, pH 7.5, 2 mM EDTA, 2 mM DTT, 0.5 M NaCl, 0.1% Nonidet P40, 10 % glycerol) for 1 hour at room temperature; after a change of the dilution buffer the filter was stored at 4°C for a minimum of 24 hours (maximum 3 days).

Radio-labelled DNA was prepared using either ca. 300 ng genomic rat DNA or salmon testis DNA that had been denatured by boiling for 5 minutes and quenched on ice. In a total volume of 20 µl, the denatured DNA was reacted with 25-40 OD units of random 14-mers, 50 µCi [α^{32}P]ATP, 5 units Klenow polymerase, 0.5 M Tris-Cl (pH 7.6), 0.1 M $MgSO_4$, 1 mM DTT, 6 mM dGTP, dCTP and dTTP for 1 hour at room temperature. The DNA mixture was passed through a Sepharose CL6B column to remove unincorporated label. The labelled DNA was added to 2.0 ml Buffer B (30 mM HEPES, pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 2 mM dithiothreitol) and allowed to hybridize to the nitrocellulose filter overnight. The filter was washed in Buffer B until no more counts were eluted in the wash solution. Visualization was by autoradiography.

6.2.2.2 Supercoiled Plasmid Nicking Assay

Approximately 1 µg plasmid DNA was incubated with 1 µg total protein in 85 mM KCl, 20 mM MOPS (pH 7.2), 6.6 mM $MnCl_2$, 10 mM DTT, 5% (w/v) BSA, at 37°C for 1 hour in a total volume of 10 µl. The reaction was stopped with the addition of 1 µl 0.5 M EDTA (pH 8.0). The DNA was analyzed on 1% agarose gels and visualized by ethidium bromide staining.

6.2.2.3 Scintillation Proximity Assay (SPA)

Oligonucleotides corresponded to the MoMLV LTR with the sequence as follows: 5'-biotin-CAGGTGGGGTCTTTCATT-3'. The

two 3' thymidines were tritiated. This biotinylated oligo was annealed to an excess of the complementary oligo 3'-GTCCACCCCAGAAAGTAA-5'. Reactions were carried out in Eppendorf tubes with a reaction volume of 85 µl (40 µl streptavidin beads, 20 µl (approx 1.2 µg) oligonucleotide, 5 µl of appropriate buffer and 20 µl of enzyme solution). Exonuclease III assays were buffered in 20 mM Tris-Cl (pH 8.0), 3 mM $MgCl_2$. Integrase assays were buffered in 85 mM KCl, 20 mM MOPS (pH 7.2), 6.6 mM $MnCl_2$, 10 mM DTT, 5% (w/v) glycerol, 0.1 mg/ml bovine serum albumin. Before adding enzyme, the bead-oligo mixture was allowed to stand undisturbed for 15-60 minutes to allow for streptavidin-biotin binding. Assays were carried out in a scintillation counter that allowed for operator control of start-stop times. Count measurements were obtained before any enzyme was added to the reaction to establish a baseline. Time zero was designated as the moment at which the enzyme was added to the bead-oligo mixture, and count measurements were recorded at either 40 second count times at 1 minute intervals or 10 second count times at 30 second intervals

6.2.2.4 Screening Scintillation Proximity Assay

Screening SPA was done as for the SPA assay but with a 10x reduction in streptavidin beads, oligonucleotide and enzyme concentrations, in a volume of 66 µl. Oligonucleotide and enzyme were allowed to react at room temperature for 5 minutes. The reaction was stopped by the addition of EDTA

(pH 8.0) to a final concentration of 75 mM. Streptavidin beads were added and the reaction mixture allowed to stand for 10 minutes to allow for streptavidin-biotin binding before counts were measured in a scintillation counter.

6.2.2.5 Oligonucleotide Cleavage and Integration Assay

LTR cleavage and autointegration assay was done as described by Craigie *et al.* (1990). Briefly, oligonucleotides L1 and L2C, corresponding to the MoMLV LTR, were purified by polyacrylamide gel electrophoresis as described by Maniatis *et al.* (1982) with the exception that the elution buffer was 0.3 M NaAc (pH 5.2), 10 mM EDTA (pH 8.0). Oligo L1 was end-labelled with [α-^{32}P]ATP with T4 polynucleotide kinase and annealed to a 4x excess of oligo L2C (oligo sequences are listed in Table 2.1). In a total volume of 15 µl, approximately 5 pmol MBP-IN fusion protein was incubated with 0.5 pmol LTR substrate in 85 mM KCl, 20 mM Mops (pH 7.2), 6.6 mM $MnCl_2$ (or $MgCl_2$ or EDTA pH 8.0), 10 mM DTT, 20% (w/v) glycerol, 100 µg/ml BSA. Reactions were incubated for 1 hour at 30°C and stopped with 15 µl 95% (v/v) formamide, 20 mM EDTA (pH 8.0), 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol. Analysis was by 20% polyacrylamide gel electrophoresis. Visualization was by autoradiography. This assay was also done in the buffer described by Jonsson *et al.* (1993); 30 mM Pipes (pH 6.8), 20 mM $MnCl_2$ (or $MgCl_2$), 30 mM DTT.

6.3 Results

6.3.1 Southwestern Blotting

Southwestern blotting was used to establish DNA binding by IN and was also intended to be used as a screening assay in the purification of IN as expressed from pIF. The first Southwestern was done with whole cell lysate from TG1 transformed with pIF and whole cell lysate from untransformed TG1 as a control. The blot (figure 6.1) shows a protein of approximately 45 kDa that co-migrates with IN as identified on Coomassie-stained SDS-PAGE gels. Although the negative control should have been TG1 transformed with pTZ19R, the results do suggest that IN as expressed from plasmid pIF binds to non-sequence specific DNA. However, this binding was not reproducible. Table 6.1 lists the Southwestern blots performed and the experimental changes made in an effort to yield positive binding.

6.3.2 Super-coiled Plasmid Nicking

IN activity was tested in a super-coiled plasmid nicking assay. Several preparations of IN as expressed from plasmid pIF were assayed for activity including protein solubilized by 6M guanidine and dialyzed against 8M urea and diluted, were assayed for activity. No conversion of supercoiled plasmid to Form II (nicked circle) was detected (data not shown). This assay was also used to screen the MBP-IN

fusion protein fractions after amylose column chromatography (section 5.3.3) (figure 6.2). The results suggest that in fractions 2 and 3 of both MBP and MBP-IN an interaction with the DNA has taken place. Whether this is actual conversion of the supercoiled plasmid to nicked circle or retardation of the DNA migration by protein interactions with the DNA is difficult to determine without further investigation. Either a bacterial nuclease has co-purified with MBP and MBP-IN or MBP possesses DNA-binding properties.

6.3.3 Oligonucleotide Auto-integration Assay

IN activity was also tested by the oligonucleotide endonuclease and auto-integration assay developed by Craigie *et al.* (1991). The assay was first carried out with purified MBP and MBP-IN (data not shown). Generalized degradation of the oligonucleotide was seen, suggesting nuclease contamination or non-specific degradation of the oligo substrate, as was reported by Jonsson *et al.* (1993); no strand transfer was detected. The assay was repeated using nuclease free BSA in the buffer with no degradation, 3' processing or strand transfer activity detected (figure 6.3). As the protein preparations had been frozen, the purification of MBP and MBP-IN was repeated on a large scale under reducing conditions (section 5.2.2.2) and the assay performed on fresh fusion protein. Also included in the assay was the soluble protein from the renaturation of IN expressed from pIF. Again, generalized degradation was

observed, even in the no protein control; no auto-integration events were seen (data not shown).

6.3.4 Scintillation Proximity Assay (SPA)

SPA was developed to examine the kinetics of the endonuclease reaction of IN. To test the validity of the assay the kinetics of Exonuclease III (Exo III), an enzyme that cleaves blunt-ended dsDNA, were investigated. Initial experiments determined the conditions under which linear rates were observed. Linear rates were obtained with an enzyme concentration of 0.2 U to 0.6 U (figure 6.4). The plot of rate versus units of enzyme (figure 6.5) shows that the decrease in counts was directly proportional to enzyme concentration demonstrating that initial rates were being measeured and that the assay system was valid for the kinetic measurement of the endonuclease reaction. Throughout these experiments the "no enzyme" controls exhibited an increasing count rate. This indicated that the beads were settling under gravity which caused the radio-labelled DNA to come into close proximity to more than one bead. This settling rate was taken into account in the rate calculations.

A stop-time screening assay was also developed in order to screen samples for activity more quickly. Although this assay was not meant to be quantitative, rates were compared for Exo III in both assays and found to be similar (data not shown).

Several preparations of IN were assayed for activity by the longer method. Enzyme assays of preparations of IN, soluble after sonication as expressed from plasmid pIF, and an identically prepared sample of *E. coli* transformed with pTZ19R, revealed the activity of a bacterial nuclease (data not shown). Samples of IN following denaturation in guanidine, dialysis against 8M urea and dilution (section 4.2.1) showed no evidence of endonuclease activity (data not shown).

Purified samples of MBP and MBP-IN were screened for activity in the screening assay, but each showed a similar drop in radioactivity. Several experiments were conducted with MBP and MBP-IN in the kinetic assay. The levels of activity for each protein were roughly equivalent. The results suggest that, although low activity was measured, activity cannot be attributed to the endonuclease reaction of IN.

Table 6.1 Listing of the samples used in Southwestern blots and changes made in the method. Experiments 1-6 were from expression in TG1. All experiments included the appropriate vector-only controls.

Expt.	Sample	Assay Conditions
1.	2,4,6,8M urea soluble and insoluble.	Hybridization 2 hours.
2.	4,6,8M urea and whole cell lysate.	Hybridization overnight (and hereafter). Incorporation of ^{32}P confirmed. Blocking buffer freshly made.
3.	6M guanidine solubilized and dialyzed in 8M urea.	Same as 2.
4.	Whole cell lysate and samples from 3.	Switch to salmon testis DNA (and hereafter). Labelled MoMLV LTR oligo as probe.
5.	32°C sonicate soluble and insoluble, whole cell lysate.	Blocking buffer switched to Blotto (and hereafter).
6.	Whole cell lysate, 8M urea insoluble, 6M guanidine soluble, 0.1% SDS soluble.	Eliminated denaturation step (and hereafter).
7.	CAG629 4,8M urea insoluble, 8M urea soluble.	Same as 6.
8.	MBP and MBP-IN soluble.	Same as 6.
9.	MBP and MBP-IN soluble.	Reducing conditions of zinc blot (section 7.2.2).

Figure 6.2 Southwestern blot of IN expressed in *E. coli* from plasmid pIF. Bacterial cells were grown and induced as described in section 2.2.21. Whole cell lysate was resuspended in reducing SDS-PAGE sample buffer and separated on a 7.5% gel. Southwestern blotting was carried out as described in section 6.2.2.1 with genomic rat DNA. Visualization was by autoradiography.

Lane 1 TG1 whole cell lysate, 10 µl loaded onto gel
Lane 2 TG1 whole cell lysate, 30 µl loaded onto gel
Lane 3 TG1 transformed with pIF, 10 µl loaded onto gel
Lane 4 TG1 transformed with pIF, 30 µl loaded onto gel

The arrow indicates the position of IN.

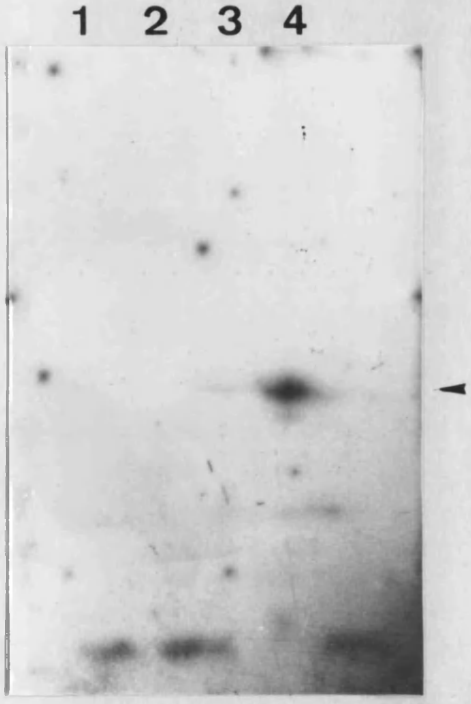

1 2 3 4

Figure 6.3 Super-coiled plasmid nicking assay of purifed MB and MBP-IN on the 1 cm^3 amylose column. The assay was done as described in section 6.2.2.2 with pRIT-5 as the plasmid substrate. Fractions 1-9 from the chromatography were screened. A. MBP fractions. B. MBP-IN fractions.

In A and B

Lane C No enzyme control

Lane 1 Fraction 1

Lane 2 Fraction 2

Lane 3 Fraction 3

Lane 4 Fraction 4

Lane 5 Fraction 5

Lane 6 Fraction 6

Lane 7 Fraction 7

Lane 8 Fraction 8

Lane 9 Fraction 9

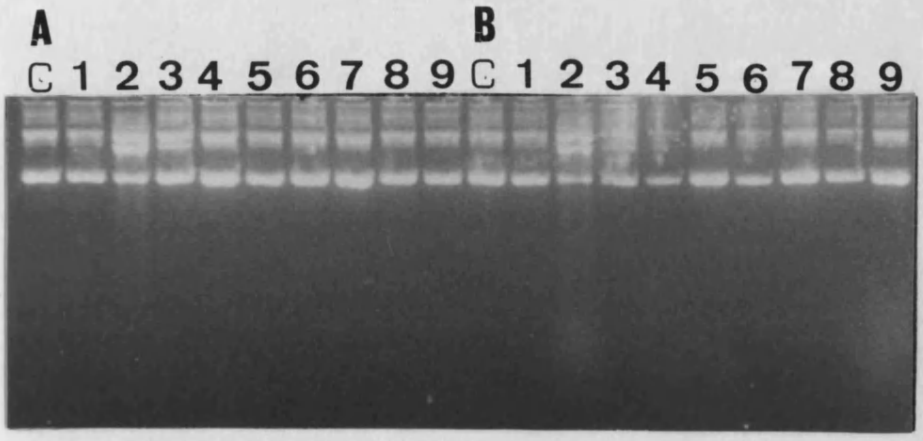

A
B
C 1 2 3 4 5 6 7 8 9 C 1 2 3 4 5 6 7 8 9

Figure 6.4 Oligonucleotide cleavage and autointegration assay. MBP and MBP-IN were purified from a 1 cm^3 amylose column as described in section 5.2.2.2. Fraction 2 from each was screened in the assay in the buffer described by Craigie *et al.* (1990) in either Mn^{2+}, Mg^{2+}, or EDTA.

Lane 1 Oligo only

Lane 2 Oligo in Mn^{2+} buffer

Lane 3 MBP in Mn^{2+}

Lane 4 MBP in Mg^{2+}

Lane 5 MBP in EDTA

Lane 6 MBP-IN in Mn^{2+}

Lane 7 MBP-IN in Mg^{2+}

Lane 8 MBP-IN in EDTA

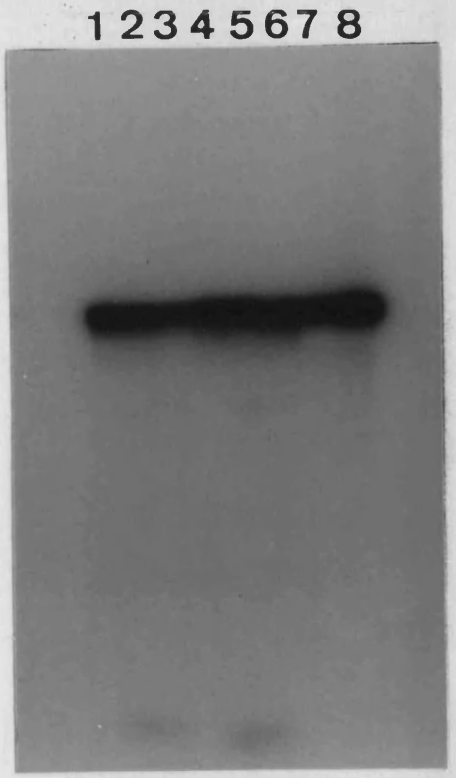

1 2 3 4 5 6 7 8

Figure 6.5 SPA assay with Exonuclease III. Biotinylated, tritiated oligonucleotides were incubated with various concentrations of Exonuclease III and the decrease in radioactivity measured by scintillation counting. Counts per minute were normalized to percent initial counts. (Not that the data for 0.1 U are not shown.)

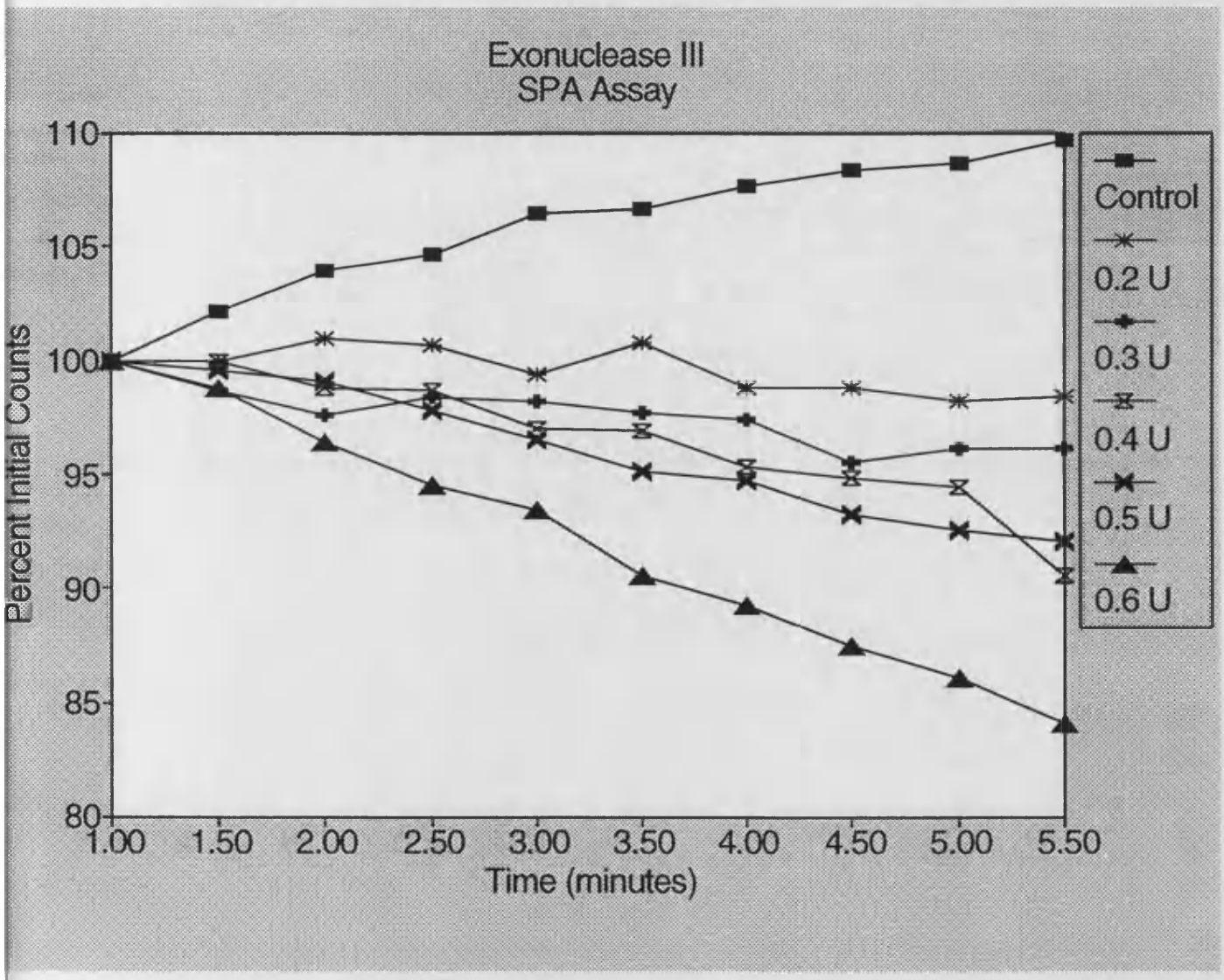
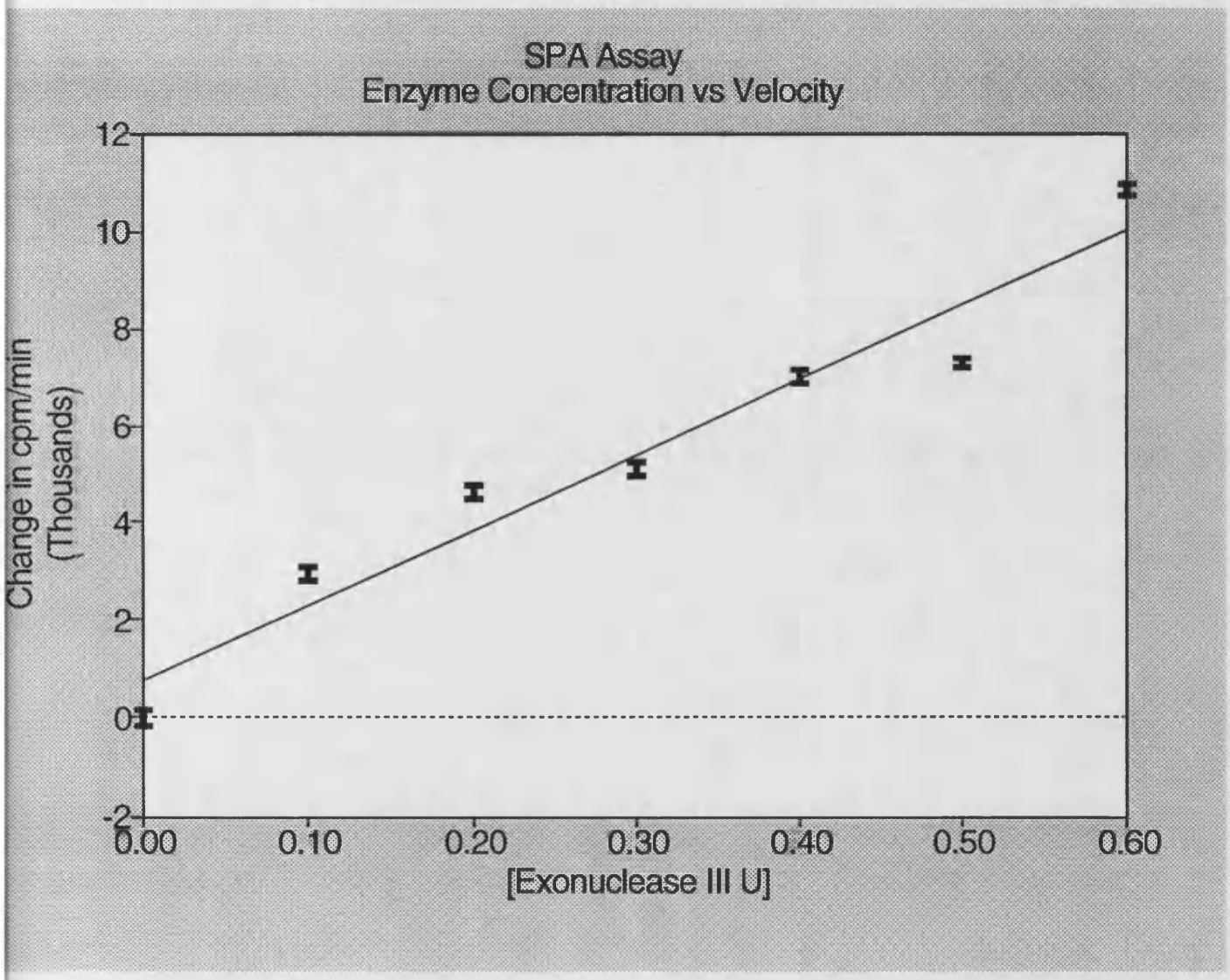

Exonuclease III
SPA Assay
Percent Initial Counts
110
105
100
95
90
85
80
1.00
1.50
2.00
2.50
3.00
3.50
4.00
4.50
5.00
5.50
Time (minutes)
Control
0.2 U
0.3 U
0.4 U
0.5 U
0.6 U

Figure 6.6 Plot of velocity versus enzyme concentration. Velocity was calculated by linear regression of the data presented in figure 6.4. Error bars reflect standard error of the mean.


SPA Assay
Enzyme Concentration vs Velocity
Change in cpm/min
(Thousands)
12
10
8
6
4
2
0
-2
0.00
0.10
0.20
0.30
0.40
0.50
0.60
[Exonuclease III U]

6.4 Discussion

DNA binding by MoMLV IN was investigated by Southwestern blotting. An initial demonstration of DNA binding by a 45 kDa protein in the cell lysate from *E. coli* transformed with pIF suggested that the recombinant IN bound DNA. Many efforts were made to repeat that result with various preparations of IN and changes made to the experimental protocol. As some of these preparations were proteins solubililized in urea or guanidine, whole cell lysates were included to rule out modification of the protein by urea or guanidine. The guanidine denaturation step was also eliminated in the later experiments for the same reason. The change in blocking buffer did result in a lower background radiation on the blots. Finally, stringent reduction conditions during renaturation were investigated.

Why these experiments failed to demonstrate that IN binds DNA may possibly be a failure of IN to refold properly following denaturation. A hypothesis that the renaturation buffer was not correct is difficult to reconcile with reports in which DNA binding is demonstrated after renaturation in that same buffer (Roth *et al.*, 1988; Krogstad and Champoux, 1990). It remains a possibility that renaturation conditions for IN may be dependent on the bacterial strain in which it was expressed; IN was expressed from pIF in *E. coli* DH5α by Krogstad and Champoux (1990), while in this study expression was from *E. coli* TG1.

Krogstad and Champoux do report variable levels of renaturation in their Southwestern assays.

Endonuclease activity was investigated by a super-coiled plasmid nicking assay. The assay failed to demonstrate IN activity; investigations were not pursued in favour of the more specific oligonucleotide autointegration and SPA assays. One interesting observation was made in the screening of column fractions of MBP and MBP-IN. Fractions 2 and 3 in both the MBP and MBP-IN purifications showed an interaction with the plasmid DNA by retarding their migration on the gel or in conversion of super-coiled DNA to nicked circle. Fractions 2 and 3 correspond to the identifed MBP and MBP-IN on Coomassie-stained SDS-PAGE gels and immunoblots. Either a bacterial DNA-binding protein or nuclease has co-purified with MBP and MBP-IN or MBP is a DNA binding protein. There is no evidence to suggest that MBP binds DNA (Kellerman and Ferenci, 1982; Duplay *et al.*, 1984).

IN preparations from the two expression systems also displayed endonuclease activity in the SPA assay; however, similar levels of activity were observed in the vector-only control. Again, this points to co-purification of a bacterial nuclease. That a bacterial nuclease can co-purify with IN expressed in *E. coli* has been demonstrated (Panet and Baltimore, 1987). IN activity was also tested by the oligonucleotide autointegration assay. Although degradation of the oligonucleotide substrate was observed, there was no evidence to suggest that 3' processing or strand transfer

had occurred. That the degradation was non-specific is suggested by the degradation observed in the "no enzyme" control.

There is every reason to believe that IN should be active as a fusion with MBP. For example, genes *hisD* and *hisC* of *Salmonella typhimurium* were fused in a spontaneous mutation. The fusion displayed catalytic activites, though not identical to the wild-type unfused proteins (Yourno *et al.*, 1970). A β-galactosidase and galactose dehydrogenase fusion also demonstrated catalytic activites. In this case β-galactosidase activity was 90-100% wild-type while galactose dehydrogenase was 40-50% wild-type (cited by Bülow, 1990). The most convincing evidence that MoMLV MBP-IN should be active is the successful demonstration of activity for an HIV-1 MBP-IN (Vink *et al.*, 1993) and avian MBP-IN (Kitamura *et al.*, 1992). That no activity was detected could be attributed to one or both of two factors; first, the MBP-IN carries two mutations that may have affected activity (see Chapter 5 discussion) and second, even though MBP-IN was soluble, the fact that IN precipitated out of solution following cleavage with Factor Xa suggests that IN may not have been folded properly.

The SPA assay was developed to examine the kinetics of IN activity. SPA was characterized with Exonuclease III; initial rates were measured and it was demonstrated that velocity was linearly proportional to enzyme concentration. These demonstrations show that the SPA could be used to examine the kinetics of IN as well as the kinetics of other

endo- and exonucleases. The assay could also have been developed to examine the kinetics of strand transfer and disintegration. By using site-directed mutagenesis catalytic residues could have been identified as well as the sequence preferences of IN by using a variety of mutant oligonucleotide sequences.

6.5 Conclusions

A variety of assays were used to demonstrate IN activity. One Southwestern blot showed DNA binding by IN. In all other assays, IN was inactive. A bacterial nuclease co-purified with IN in both expression systems. A continuous direct kinetic assay was developed and characterized using Exonuclease III and could be used to examine the kinetics of IN once the problems of solubility have been overcome.

Chapter 7

This chapter describes the demonstration of zinc binding by MoMLV IN and the construction of site-specific mutants to examine the residues involved in zinc binding.

7.1 Introduction

A comparison of 5 retroviral IN sequences by Johnson and co-workers (1986) identified a putative N-terminal zinc finger motif. This was later confirmed by Khan *et al.* (1991) in a comparison of 80 retroviral and retrotransposon sequences. The zinc finger motif was first described by Miller *et al.* (1985) from Trascription Factor (TF) IIIA and has subsequently been identified in numerous proteins [reviewed by Evans and Hollenberg (1988), Berg (1990), and Coleman (1992)].The zinc finger is so named for the co-ordination of zinc to four cysteine or histidine residues arranged in a particular manner, usually -Cys-X_{2-4}-Cys-X_{12-13}-His-X_{3-4}-His-, where X is any other amino acid. This motif is implicated in nucleic acid recognition and binding, usually to dsDNA, less commonly to dsRNA, ssDNA, and ssRNA.

The zinc finger of the retroviral integrases differs in three significant ways from the classical zinc finger: i) the sequence of zinc co-ordinating residues is HHCC rather than CCHH, ii) the loop between the 2nd and 3rd zinc co-ordinating residues is twice as long (HIV-1) or three times as long (MoMLV), iii) it is present in a single copy, not a

string of 3-12 similar motifs. Similarities include a role in DNA binding and in structure. The crystal structures of a TFIIIA zinc finger (Berg, 1988) and Zif268, a mouse immediate early protein, (Pavletich and Pabo, 1991) have been solved and show the histidine region in an α-helix and the cysteine region in an anti-parallel β-sheet. The zinc finger of HIV-1 (residues 1-55) has been studied by spectroscopy and reveals the same arrangement of α-helix and β-sheet (Burke *et al.*, 1992).

The expected role of the putative zinc finger of IN is sequence-specific DNA binding. However, there is conclusive evidence that DNA binding activity is localized in the C-terminus of HIV-1 (Schauer and Billich, 1992; Vink *et al.* 1993) and that the HHCC region of MoMLV IN and avian IN is not needed for DNA binding (Donehower, 1988; Roth *et al.*, 1989, 1990; Khan *et al.*, 1991) Analysis of mutations of the zinc finger suggest possible roles: specific interactions with the LTR for avian IN (Khan *et al.*, 1991), 3' processing for HIV-1 IN (LaFemina *et al.*, 1992; Vink *et al.*, 1993), 3' processing and strand transfer for HIV-1 IN (Leavitt *et al.*, 1993; Vincent *et al.*, 1993) and MoMLV (Jonsson and Roth, 1993), binding and positioning of the LTR ends for HIV-2(van Gent *et al.*, 1992), and protein/protein interactions for HIV-1 (Engelman and Craigie, 1992). The observations that DNA binding is associated with the C-terminus, and that the HHCC region is not needed for DNA binding are not incompatible with a role of the HHCC region in LTR binding. IN acts on two DNA molecules: the viral LTR and the target

DNA. It follows that there may be two DNA binding centres in the protein: the HHCC region for LTR binding and a region in the C-terminus for target DNA binding.

7.2 Materials and Methods

7.2.1 Materials

Taq polymerase for overlap extension was from Perkin Elmer. ^{65}Zn was from NEN Dupont.

7.2.2 Methods

7.2.2.1 PCR Overlap Extension

The two DNA fragments were amplified by PCR as described in section 5.2.2.1 and agarose gel purified by freeze-squeeze (section 2.2.14). The purified fragments were placed in the usual PCR reaction mix without end primers, melted for 5 minutes at 94°C and cycled for 5 cycles at 94°C for 1 minute, 37°C for 1 minute for annealing and 72°C for 2 minutes for extension. End primers were added (50 pmol, a 2.5x excess) and cycled for 33 cycles as usual. Analysis of products was by agarose gel electrophoresis (section 2.2.1).

7.2.2.2 Zinc Blotting

Zinc blotting was by the method of Barbosa *et al.* (1989). Following protein transfer to nitrocellulose, the filter was washed in Renaturation Buffer (100 mM Tris, 50 mM NaCl, 10 mM DTT; pH 6.8) for 50 minutes with three changes of buffer, followed by a 10 minute wash in 10 mM β-mercaptoethanol in Metal Binding Buffer (MBB)(100 mM Tris, 50 mM NaCl; pH 6.8). The filter was sealed in a bag and washed for 10 minutes in MBB through which N_2 had been bubbled. ^{65}Zn was added to a concentration of 10 µCi/lane and diluted with MBB to a $ZnCl_2$ concentration of 50 µM. Incubation was for 30 minutes with gentle agitation. The filter was washed for 30 minutes in 1 mM DTT in MBB with three changes of buffer. Visualization was by autoradiography

7.3 Results

7.3.1 Zinc Blotting

Samples for zinc blotting were prepared from *E. coli* transformed with plasmid pTZ19R or pIF. Insoluble fractions after detergent washing were analyzed by SDS-PAGE, immunoblotting and zinc blotting (figure 7.1). The zinc blot showed the presence of a zinc binding protein at 45 kDa that co-migrated with IN as identified by Coomassie staining and immunoblotting. This band was not present in the parent vector controls. As normally-soluble proteins may become

trapped in insoluble inclusion bodies, a whole cell lysate of *E. coli* transformed with pTZ19R was included as an additional control. No zinc-binding or immunoreactive bands were seen. The observed zinc binding activity can therefore be attributed to IN.

Zinc binding was lost if reducing conditions were not maintained during renaturation. As IN also requires Mn^{2+} for activity (Ca^{2+} and Mg^{2+} can substitue to a lesser degree) the possibility exists that $^{65}Zn^{2+}$ was binding adventitiously to the Mn^{2+} site. However, no decrease in $^{65}Zn^{2+}$ binding activity was seen in the presence of 10 mM Mn^{2+}, Ca^{2+}, or Mg^{2+} (a 200-fold molar excess over Zn^{2+}). In contrast, binding of $^{65}Zn^{2+}$ was greatly reduced in the presence of 1 mM non-labelled $ZnCl_2$.

7.3.2 Construction of pMIN-CC/AA and pMIN-CC/SS

In order to demonstrate more conclusively that the zinc was binding to the putative zinc-binding motif, two mutants were planned for construction, pMIN-CC/AA replaces both cysteines of the zinc-binding domain with alanine and pMIN-CC/SS replaces both cysteines with serine. Two considerations were important in the choice of replacement amino acids: proper folding and side chain interactions with zinc. Serine most closely resembles cysteine in structure and would be least likely to disrupt secondary structure, but its electronegative oxygen atom may interact with the zinc. Alanine may disrupt the secondary structure of the zinc

binding domain but has a side chain without zinc binding capabilities.

Site-directed mutagenesis by PCR overlap extension was employed to construct the mutants and is presented in detail in figure 7.2. After each PCR reaction or restriction digestion the DNA was purified from agarose by freeze-squeeze, extracted with phenol and ethanol precipitated. Recombinants were analyzed by restriction digestion after mini-prep of plasmid DNA. The mutations made to create pMIN-CC/AA destroyed a *Hind*III and *Hg*IAI site at positions 4894 and 4900 respectively; pMIN-CC/SS destroyed the *Hg*IAI site at 4900. pMIN-CC/AA was identifed from 11 recombinant colonies, of which one gave the correct restriction map (figure 7.3). This clone was sequenced from a maxi-preparation of DNA (LiCl) from the polylinker through the *Sph*I site, which confimed the desired muations. Also noted were the mutations previously described for pMIN (section 5.3.5).

Construction of pMIN-CC/SS proved more problematic. Of the 12 recombinants obtained from the ligation and transformation, only 3 possessed plasmids of the correct size: S1, S4, and S7. Analyses of digestion with *Hg*IAI were difficult to interpret due to the high number of *Hg*IAI restriction sites in both the parent vector and IN coding region; all did retain the *Hind*III site (data not shown). All 3 of these recombinants were instead sequenced to screen for the mutations. S1 and S7 proved to be wild-type; S4,

though difficult to sequence, bore no resemblence to the IN coding region.

A second attempt was made to construct pMIN-CC/SS. A different strategy was employed, namely the use of oligo 5191 in place of oligo Pst2 in the PCR reaction, thereby eliminating the need to digest the fragment 965 with *SfiI*. This results in a fragment size of 329 bp. Both fragments were amplified by PCR (section 5.2.2.1), purified from agarose by freeze-squeeze, extracted with phenol, and then ethanol precipitated. PCR overlap extension was unsuccessful in producing a fragment of 622 bp, despite varying the concentration of the fragments, melting and extension time in the PCR reaction, and the concentration of end primers and Taq DNA polymerase (data not shown). Due to time constraints, construction of pMIN-CC/SS was abandonded.

7.3.3 Expression and Purification of MBP-IN(Ala)

pMIN-CC/AA was transformed into *E. coli* TB1 by the $CaCl_2$ method. MBP-IN(Ala) (MBP-IN fusion carrying the cysteine to alanine mutations) was expressed exactly as described for the large scale purification of MBP-IN (section 5.2.2.2)(figure 7.4). The Coomassie stained SDS-PAGE gel and immunoblot demonstrate that MBP-IN(Ala) was expressed and purifed in a similar manner to MBP-IN. MBP-IN(Ala) also precipitated out of solution after cleavage with Factor Xa, was soluble in 4M urea, but failed to renature on dialysis as described (section 5.2.2.4)(figure 7.5).

7.3.4 Zinc Blotting with MBP-IN and MBP-IN(Ala)

Zinc binding was first demonstrated with MBP-IN (figure 7.6). The zinc blot revealed the presence of several zinc binding proteins that co-migrate with MBP-IN and degradation products (previously described). IN, as a fusion with MBP, retains the ability to bind zinc. MBP-IN(Ala) was then tested for zinc binding. Four zinc blots were performed; all were negative for zinc binding to any proteins in the samples (including IN as expressed from plasmid pIF as a positive control). There was evidence to suggest inadequate protein transfer in the first blot; this was repeated with excellent transfer the second time. In the third attempt, fresh Metal Binding Buffer and more stringent care to maintain reducing conditions were employed, and in the fourth attempt, all buffers were freshly made as well as using a different preparation of $^{65}Zn^{2+}$ with a higher specific activity. (Time constraints did not allow any further attempts to demonstrate zinc binding or lack of it.)

Figure 7.1 Zinc binding by MoMLV IN. All samples were resuspended in reducing SDS-PAGE sample buffer and separa on an 8.0% polyacrylamide gel. A. Zinc blot. B. Coomassie-stained gel. C. Immunoblot with a 1:400 dilut anti-KK28 anti-serum, 1:10,000 dilution goat anti-rabbit alkaline phosphatase conjugate. Visualization was with NBT/BCIP. Open arrow indicates the position of IN.

In A, B, and C

Lane 1 TG1 transformed with pIF, insoluble fraction after detergent wash

Lane 2 TG1 transformed with pTZ19R, insoluble fraction af detergent wash

Lane 3 TG1 transformed with pTZ19R, whole cell lysate

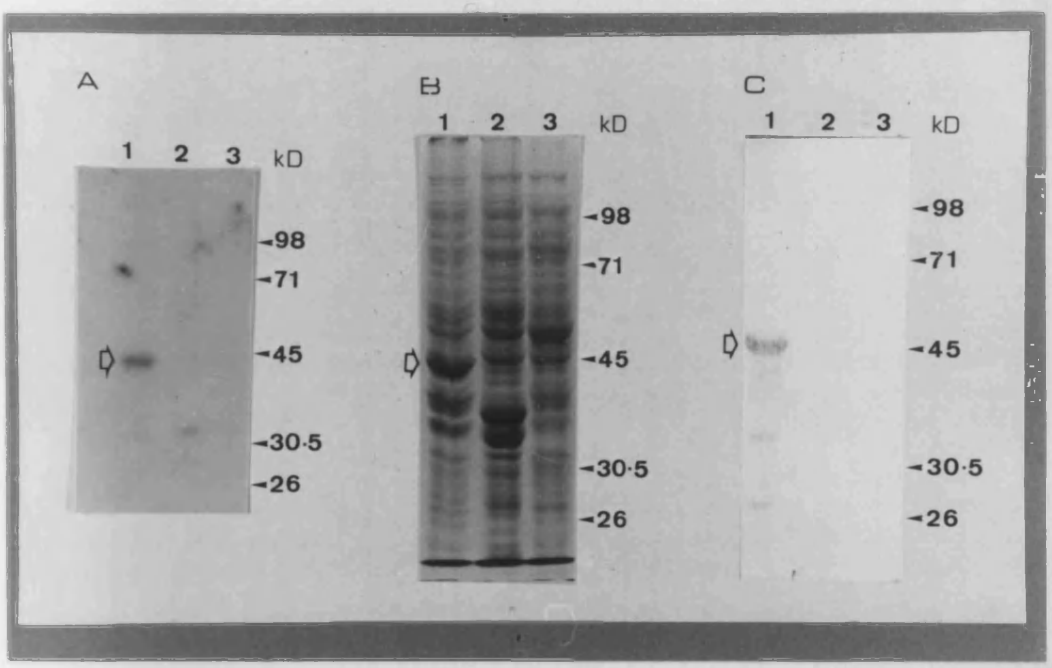

A
1 2 3 kD
98
71
45
30·5
26
B
1 2 3 kD
98
71
45
30·5
26
C
1 2 3 kD
98
71
45
30·5
26

Figure 7.2 Illustration of the construction of the mutant vectors. The first step was the PCR amplification of the two fragments. Letters indicated the primers used (sequences are listed in table 2.1). Numbers indicate the fragment sizes in base pairs. Filled circles indicate the relative position of the mutations. After each digestion PCR reaction the products were purified from agarose by freeze-squeeze, extracted with phenol and ethanol precipitated. Final purification before ligation was on a NEBsorb column.

A. Oligo Eag

B. Oligo CysAla2 or CysSer2

C. Oligo CysAla or CysSer

D. Oligo Pst2

E. Oligo 5191

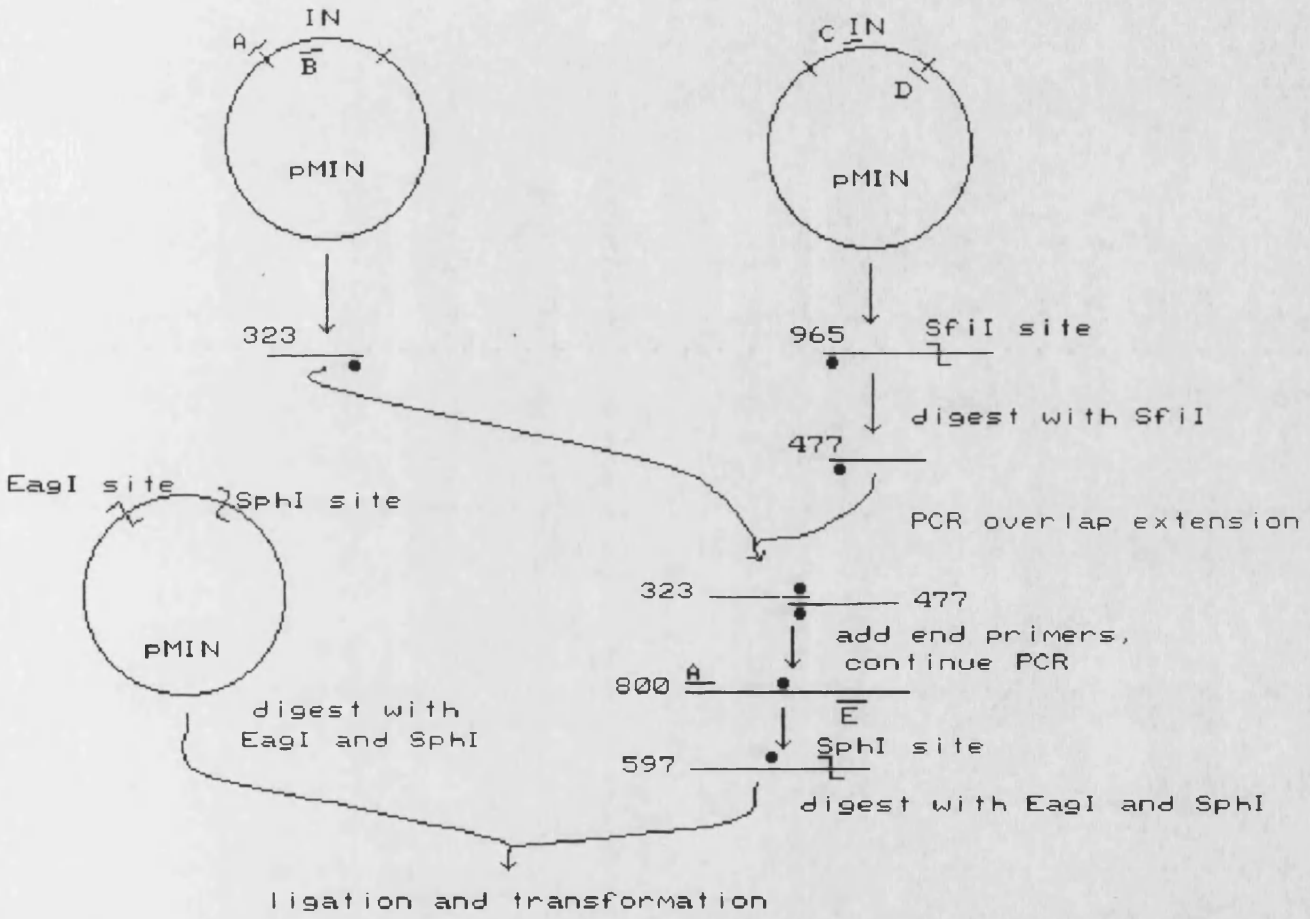

IN
A
B
pMIN
C IN
D
pMIN
323
965
SfiI site
digest with SfiI
477
EagI site
SphI site
PCR overlap extension
pMIN
323
477
add end primers,
continue PCR
800
A
B
digest with
EagI and SphI
SphI site
597
digest with EagI and SphI
ligation and transformation

Figure 7.3 Restriction mapping of pMIN-CC/AA. Plasmid DNA was prepared from a LiCl maxi-prep and digested with *Hg*IAI, *Hind*III, and *Pst*I. Analysis was on a 1.0% agarose gel; visualization was by ethidium bromide staining. The gel shows the destroyed *Hg*IAI and *Hind*III sites; the arrows indicate the position of the fragment missing in the mutant plasmid.

Lane 1 pMIN digested with *Hg*IAI

Lane 2 pMIN-CC/AA digested with *Hg*IAI

Lane 3 pMIN digested with *Hind*III

Lane 4 pMIN-CC/AA digested with *Hind*III

Lane 5 pMIN digested with *Pst*I

Lane 6 pMIN-CC/AA digested with *Pst*I

Lane 7 *Hind*III digested lambda size markers

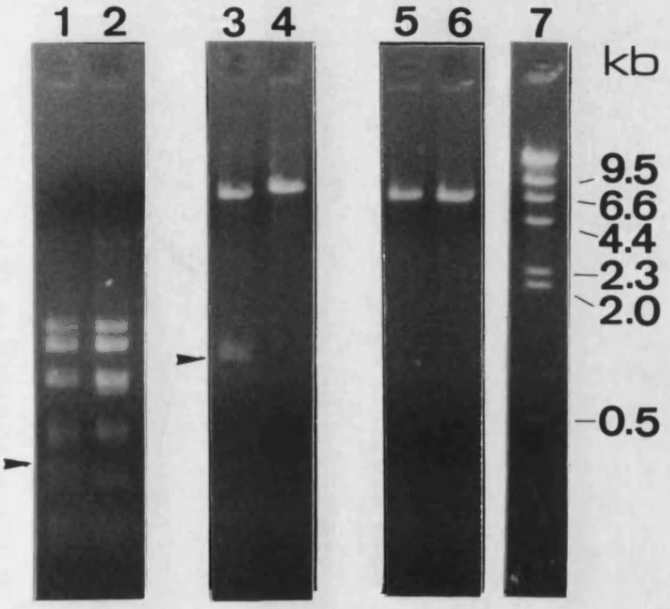

1 2
3 4
5 6
7
kb
9.5
6.6
4.4
2.3
2.0
0.5

Figure 7.4 Coomassie-stained SDS-PAGE gel and immunoblot of MBP-IN(Ala) purification. TB1 transformed with pMIN-CC/AA was grown and induced as described in section 2.2.22. The cells were lysed by sonication as described in section 2.2.23; aliquots were taken for analysis. Amylose column chromatography was as described in section 5.2.2.2 under reducing conditions. Samples were resuspended in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. A. Coomassie-stained gel. B. Immunoblot with a 1:100 dilution anti-KK28 anti-serum and a 1:10,000 dilution goat anti-rabbit IgG alkaline phosphatase. Visualization was with NBT/BCIP. The arrow indicates the position of MBP-IN(Ala). Molecular weight standards are in kDa.

In A and B

Lane 1 Insoluble fraction after sonication

Lane 2 Soluble fraction after sonication

Lane 3 Void volume

Lane 4 Wash flow through

Lane 5 Elution fraction 1

Lane 6 Elution fraction 2

Lane 7 Elution fraction 3

Lane 8 Elution fraction 4

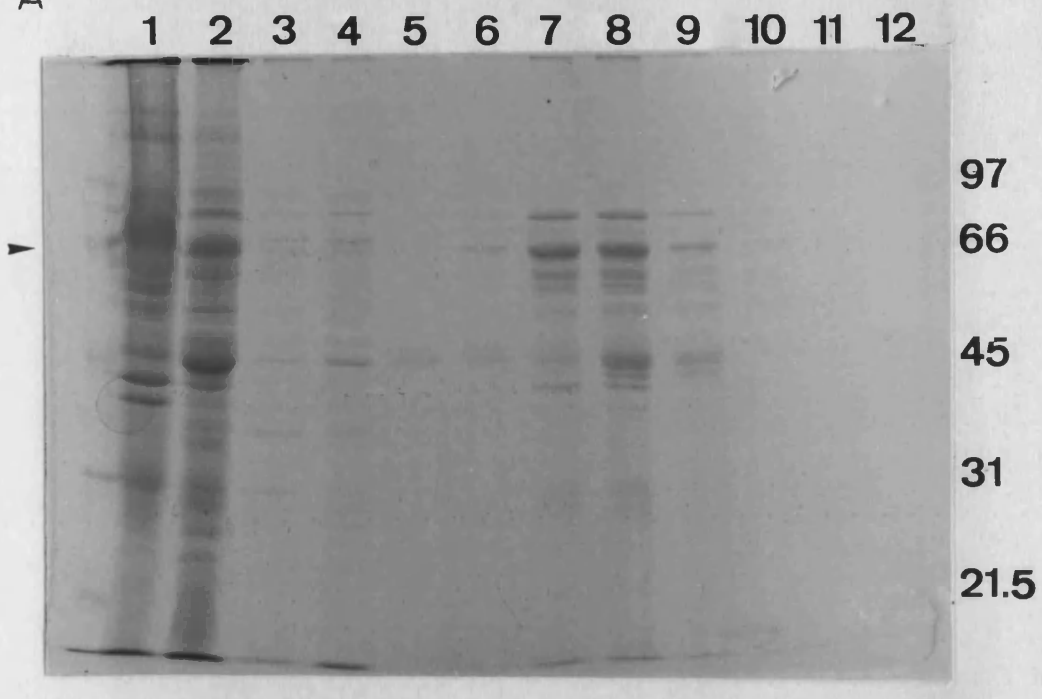
Lane 9 Elution fraction 5

Lane 10 Elution fraction 6

Lane 11 Elution fraction 7

Lane 12 Elution fraction 8

A


1 2 3 4 5 6 7 8 9 10 11 12
97
66
45
31
21.5

B

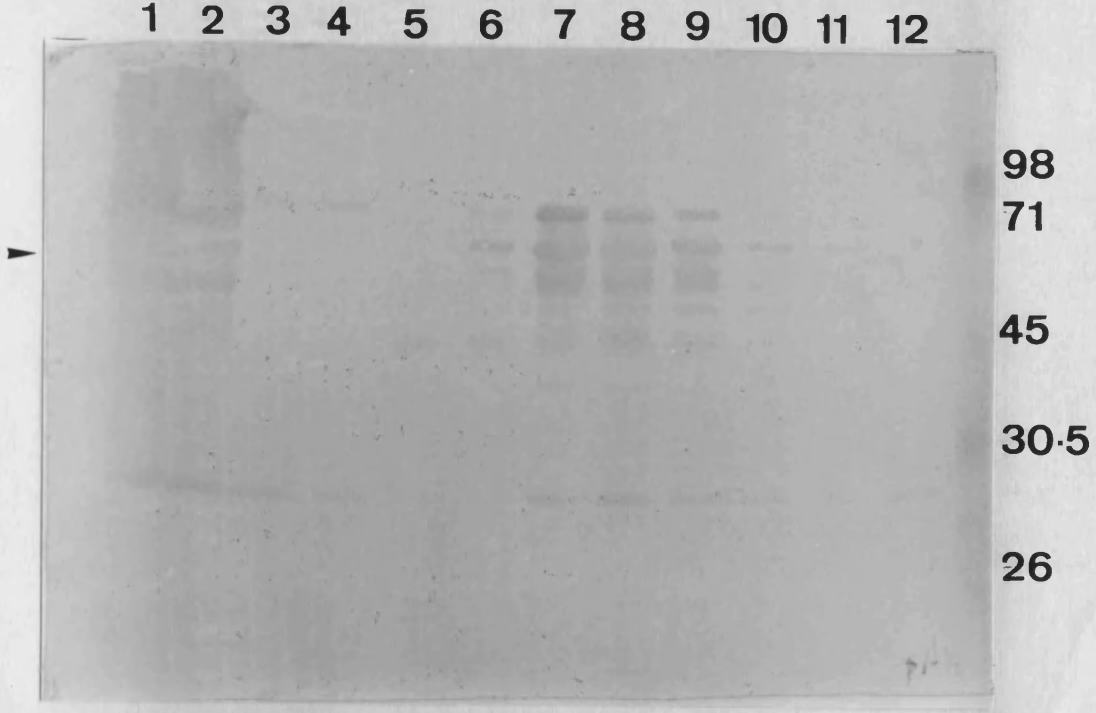

1 2 3 4 5 6 7 8 9 10 11 12
98
71
45
30·5
26

Figure 7.5 Immunoblot of MBP-IN(Ala) cleaved with Factor Xa, solubilized in 4M urea, and dialyzed as described in section 5.2.2.4. Samples were taken up in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. Immunoblotting was with a 1:100 dilution anti-KK28 anti-serum and a 1:10,000 dilution goat anti-rabbit IgG alkaline phosphatase. Visualization was with NBT/BCIP.

Lane 1 Elution fraction 4

Lane 2 Soluble fraction after cleavage

Lane 3 Soluble fraction in 4M urea

Lane 4 Insoluble fraction after dialysis

Lane 5 Soluble fraction after dialysis

Lane 6 Molecular weight standards (kDa)

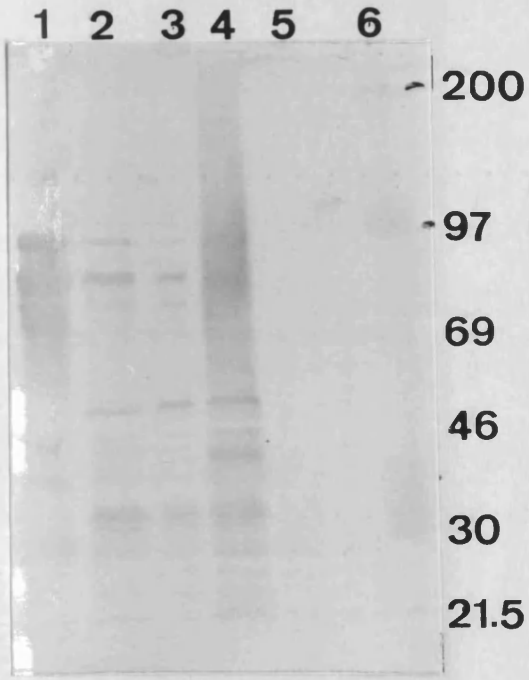
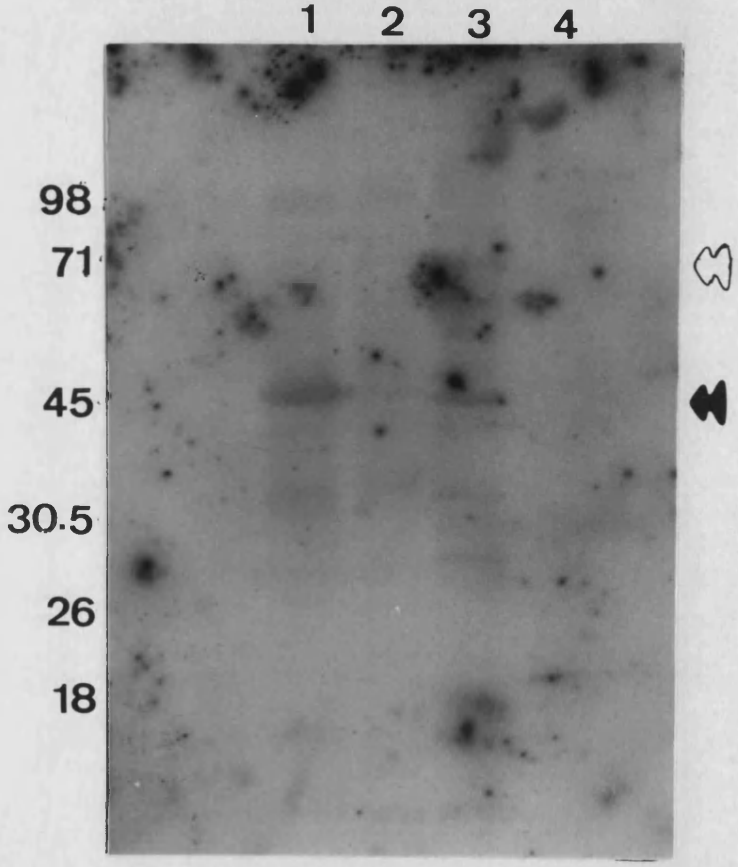

1 2 3 4 5 6
200
97
69
46
30
21.5

Figure 7.6 Zinc blot of MBP and MBP-IN. Insoluble fractions after sonication and detergent wash of TG1 transformed with either pTZ19R or pIF and soluble fractions after sonication of TB1 transformed with pMAL-c or pMIN. Samples were resuspended in reducing SDS-PAGE sample buffer and separated on a 10% polyacrylamide gel. Zinc blotting was done as described in section 7.2.2.2. The open arrow indicates the position of MBP-IN, the closed arrow the position of IN. Molecular weight standards are in kDa.

Lane 1 TG1 transformed with pIF, wash insoluble
Lane 2 TG1 transformed with pTZ19R, wash insoluble
Lane 3 TB1 transformed with pMIN, sonicate soluble
Lane 4 TB1 transformed with pMAL-c, sonicate soluble


1
2
3
4
98
71
45
30.5
26
18

7.4 Discussion

MoMLV IN as expressed from pIF was shown to bind zinc *in vitro*. This was the first demonstration of zinc binding for MoMLV IN. Basu and Varmus (1990) were successful in purifying IN expressed in yeast on a zinc affinity column which also suggests IN has an affinity for zinc. That zinc binding was oxidation-sensitive and was retained in the presence of competing divalent cations suggests that zinc is binding to a distinct site and that cysteines are involved at that site. MBP-IN also showed zinc binding, demonstrating that IN retains the ability to bind zinc as a fusion with MBP.

Site-directed mutagenesis was employed to determine if the zinc-binding site was the zinc finger motif. The mutants chosen for construction were based on maintenence of secondary structure, yet avoiding residues that could co-ordinate zinc. While this work was in progress, Bushman and co-workers (1993) reported zinc binding by HIV-1 IN. Mutations of the zinc co-ordinating residues, including a C40S/C43S double mutant, showed decreased, not abolished, zinc binding. This observation confirms the choice of the two planned mutants, as well as providing evidence that zinc is most likely binding to the zinc finger of HIV-1.

Construction of mutant pMIN-CC/AA was successfully accomplished and the mutations were confirmed by restriction mapping and sequencing. Two attempts were made to construct pMIN-CC/SS. Three recombinants of the correct size were

analyzed by sequencing. Two were wild-type which suggests inefficient digestion of the wild-type vector; the other bore no resemblance to the IN coding region and was not analyzed further. The second attempt, which utilized a different cloning strategy, was unsuccessful, failing in the PCR overlap extension. This was possibly due to the fragments being almost equal in length.

MBP-IN(Ala) was expressed in *E. coli* TB1 and purified under reducing conditions as described in section 5.3.4. Some HIV-1 IN proteins with mutations in the zinc co-ordinating residues failed to be expressed or were insoluble (LaFemina *et al.*, 1992) or failed to purify (Leavitt *et al.*, 1993). Although soluble when expresssed as a fusion, MBP-IN(Ala) precipitated out of solution following cleavage with Factor Xa. This finding is consistent with the hypothesis that proline content is the most important factor in improper folding of IN and not incorrect disulphide bond formation. MoMLV IN has three cysteines and two of these were mutated to alanine in MPB-IN(Ala) yet it showed no greater solubility than MBP-IN.

MBP-IN(Ala) was tested for activity in the oligonucleotide autointegration assay in conjunction with MBP-IN. MBP-IN(Ala) showed no 3' processing or strand transfer (data not shown). Whether this was due to the mutations or to inactivity due to improper folding cannot be determined. MBP-IN(Ala) was also tested for zinc binding, but consistent results were not obtained. Efforts were made to discover the conditions responsible for the failure of

the assay, including maintenance of stricter reducing conditions and using a $^{65}Zn^{2+}$ preparation with higher specific activity. Perhaps substituting β-mercaptoethanol for DTT throughout the assay may have proven successful.

The demonstration that MoMLV binds zinc supports the hypothesis that the zinc finger motif of IN is functioning in a similar manner to the sequence-specific DNA binding of classical zinc fingers. Jonsson and Roth (1993) showed that chemical modification of the cysteines of MoMLV IN resulted in a loss of 3' processing and strand transfer. Evidence from studies of HIV-1, HIV-2 and avian IN also indicate a role in 3' processing and strand transfer for the zinc finger domain (Khan *et al.*, 1991; Engelman and Craigie, 1992; van Gent *et al.*, 1992; Leavitt *et al.*, 1993; Vincent *et al.*, 1993). 3' processing involves the recognition of a specific DNA sequence. That strand transfer is affected as well is not surprising since 3' processing is a necessary prerequisite for strand transfer (Craigie *et al.*, 1990). The integrity of the zinc finger domain is necessary for viral infectivity *in vivo* (Donehower, 1988; Roth *et al.*, 1989).

7.5 Conclusions

Zinc binding was demonstrated by IN as expressed from both expression systems. A distinct zinc binding site involving cysteine residues was postulated. Two mutants were planned to examine the residues involved in zinc binding; one of

these was successfully constructed. Zinc affinity, however, was not determined for that mutant. The demonstration of zinc binding suggests the zinc finger may function in sequence-specific DNA recognition of the LTR ends.

Chapter 8

8.0 Discussion

The IN protein of MoMLV was investigated in two expression systems, in a variety of activity assays, and in zinc binding. The first expression system, as a fusion with the first three amino acids of β-galactosidase, expressed IN in insoluble inclusion bodies. IN was denatured in urea and guanidine and renatured in a variety of conditions. Analysis of these experiments by Coomassie-stained SDS-PAGE gels was complicated by a bacterial protein that co-migrated with IN and displayed similar solubility characteristics. For this reason two synthetic peptides were synthesized and used to produce rabbit anti-sera. Both sera were capable of recognizing IN on immunoblots, although anti-KK28 was more efficient than anti-KK334. The anti-sera were helpful in determining the nature of degradation products in the expression systems, as well as identifying soluble IN following denaturation and refolding that might otherwise have been missed.

The second expression system, as a fusion with MBP, expressed MBP-IN in a soluble form, but following cleavage of the fusion with Factor Xa the IN portion precipitated out of solution. This finding suggested that IN was not folded properly, even though soluble as a fusion protein. Inappropriate disulphide bond formation is unlikely to be the only cause of improper folding as mutation of two of

three cysteines of IN did not improve solubility. The hypothesis was put forward that a 7.8% proline content may account for the insoluble nature of IN.

IN from both expression systems was assayed *in vitro* for activity. The assays revealed that IN was expressed in an inactive form. A bacterial nuclease co-purified with MBP-IN. A novel, direct, continuous kinetic assay was developed to measure endonuclease kinetics. Should soluble, active IN be produced, this assay would be a valuable tool in characterizing the enzyme.

Zinc binding by IN was demonstrated. One site-specific mutant of the putative zinc co-ordinating residues was constructed, but was not successful in identifying the zinc co-ordinating residues.

To date, there is only one report in the literature that describes the successful expression of MoMLV IN in *E. coli* (Jonsson *et al.*, 1993)(hereafter referred to as IN$(His)_6$). The strategy of Jonsson was to delete the 28 C-terminal amino acids (giving a proline content decrease to 7.6%) that had previously been shown not to be required for activity (Roth, 1988). A hexahistidine tail was added for the purpose of purification on a nickel affinity column. The final yield of active IN was only 1.0% of the recombinant expressed, a finding which highlights the difficulty of expressing IN in *E. coli*. Jonsson and co-workers found an absolute requirement for the maintenance of reducing conditions and very slow renaturation. These considerations were taken into account in the purification

and renaturation of MBP-IN and MBP-IN(Ala), although the procedures were not successful. Reduction was felt to be necessary for proper folding of the zinc-binding domain, a finding in agreement with the observation of Párraga *et al.* (1990) that zinc binding is necessary for α-helix formation and domain folding in a zinc finger from yeast ADR1, and also in agreement with the *in vitro* zinc-binding assay described in Chapter 7.

$IN(His)_6$ is not the native sequence and this is reflected in various differences in reaction conditions. For instance, $IN(His)_6$ was inactive when Mg^{2+} was substituted for Mn^{2+}. Both IN as expressed in baculovirus (Bushman *et al.*, 1990) and IN purified from disrupted virions (Ishimoto *et al.*, 1991) were active *in vitro* with Mg^{2+} as the divalent cation. The *in vitro* reaction buffer also differs. Craigie *et al.* (1990) report an absolute requirement for a small basic protein, such as BSA, for demonstration of IN activity (expressed from baculovirus) *in vitro*; BSA is not required by $IN(His)_6$. Whether $IN(His)_6$ accurately mimics IN activities is open to question.

Why are HIV IN and avian IN successfully expressed in *E. coli* and not MoMLV IN? First, the amino acid composition is different; HIV-1 has only 10 prolines and second, MoMLV IN has a 36 amino acid insertion near the C-terminus which results in a protein 14 kDa larger than HIV IN or avian IN. Why is MoMLV IN successfully expressed from baculovirus and not *E. coli*? As baculovirus expression is a eukaryotic system and MoMLV infects eukaryotes, the possibility exists

that MoMLV IN requires eukaryotic cellular factors for proper folding, perhaps PPI or chaperones. Post-translational modification remains a possibility; however, there is no evidence that IN is glycosylated or phosphorylated (Tanese *et al.*, 1986). It was not until the discovery that HIV-1 was soluble and active when expressed in *E. coli* BL21 (Sherman and Fyfe, 1990) that significant progress was made in the characterization of that enzyme. That such a breakthrough should occur for MoMLV IN seems unlikely.

Future work on the characterization of MoMLV IN should perhaps concentrate on baculovirus expression or IN from disrupted virions. Certainly, MoMLV IN would be an excellent model protein for studies of *in vitro* and *in vivo* folding. With soluble, active IN, the enzyme would be characterized using the SPA assay: both the 3' processing SPA and a strand transfer SPA which would be developed. Using site-directed mutagenesis, active and catalytic residues would be identified. SPA could also be developed as a screen for inhibitors of IN. IN could be purified to homogeneity, crystallized, and the 3-D structure determined by X-ray crystallography. With the structure and the information about active and catalytic residues, rational inhibitors could be designed. The results from these studies could also be used to compare and contrast MoMLV IN with HIV-1 and 2, avian, and other retroviral integrases that might give evidence to evolutionary origins and relationships of the retroviruses.

References

Adachi, A., Ono, N., Sakai, H., Ogawa, K., Shibata, R., Kiyomasu, T., Masuike, H., Ueda, S., (1991). Generation and characteriztion of the human immunodeficiency virus type 1 mutants. *Archives of Virology* **117**, 45-58.

Albritton, L.M., Tseng, L., Scadden, D., Cunningham, J.M., (1989). A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. *Cell* **57**, 659-666.

Atherton, E., Sheppard, R.C., (1985). solid phase peptide synthesis using N_{α}-Fluorenylmethoxycarbonylamino acid pentafluorophyl esters. *Journal of the Chemical Society Chemical Communications* ?, 165-166.

Atherton, E., Sheppard, R.C., (1989). Solid Phase Peptide Synthesis: A Practical Approach. IRL Press at Oxford University Press, Oxford.

Baker, T.A., Grossman, A.D., Gross, C.A., (1984). A gene regulating the heat shock response in *Esherichia coli* also affect proteolysis. *Proceedings of the National Academy of Sciences USA* **81**, 6779-6783.

Baltimore, D., (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209-1211.

Barbosa, M.S., Lowy, D.R., Schiller, J.T., (1989). Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. *Journal of Virology* **63**, 1404-1407.

Basu, S., Varmus, H.E., (1990). Moloney murine leukemia virus integration protein produced in yeast binds specifically to viral *att* sites. *Journal of Virology* **64**, 5617-5625.

Berg, J.M., (1990). Zinc finger domains: Hypotheses and current knowledge. *Annual Reviews in Biophysics and Biophysical Chemistry* **19**, 405-421.

Billich, A., Schauer, M., Franks, S., Rosenwirth, B., Billich, S., (1992). HIV-1 integrase: High-level production and screening assay for the endonucleolytic activity. *Antiviral Chemistry and Chemotherapy* **3**, 113-119.

Boeke, J.D., (1988). Retrotransposons. p. 59-103. *In* E. Domingo, J.J., Holland, P. Ahlquist (eds.), RNA Genetics Vol. II Retroviruses, Viroids, and RNA Recombination. CRC Press, Boca Raton, Florida.

Boeke, J.D., Chapman, K.B., (1991). Retrotransposition mechanisms. *Current Opinion in Cell Biology* **3**, 502-507.

Botterman, J., Zabeau, M., (1985). High-level production of the *Eco*RI endonuclease under the control of the p_L promoter of bacteriophage lambda. *Gene* **37**, 229-239.

Bowerman, B., Brown, P.O., Bishop, J.M., Varmus, H.E., (1989). A nucleoprotein complex mediates the integration of retroviral DNA. *Genes and Development* **3**, 469-478.

Broderson, J.R., (1989). A retrospective review of lesions associated with the use of Freund's adjuvant. *Laboratory Animal Science* **39**, 400-405.

Brown, P.O., (1990). Integration of Retroviral DNA. p. 19-48. *In* R. Swanstrom and P.K. Vogt (ed.), Current topics in Microbiology and Immunology: Retroviruses, Strategies of Replication. Vol. 157. Springer-Verlag, Berlin.

Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M., (1989). Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proceedings of the National Academy of Sciences USA* **86**, 2525-2529.

Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M., (1987). Correct integration of retroviral DNA *in vitro*. *Cell* **49**, 347-356.

Bukrinsky, M.I., Haggerty, S., Dempsey, M.P., Sharova, N., Adzhubei, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., Stevenson, M., (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of dividing cells. *Nature* **365**, 666-669.

Bülow, L, (1990). Preparation of artificial bifunctional enzymes by gene fusion. *Biochemical Society Symposium* **57**, 123-133.

Burke, C.J., Sanyal, G., Bruner, M.W., Ryan, J.A., LaFemina, R.L., Robbins, H.L., Zeft, A.S., Middaugh, C.R., Cordingly, M.G., (1992). Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *Journal of Biological Chemistry* **267**, 9639-9644.

Bushman, F.D., Craigie, R., (1990). Sequence requirements for integration of Moloney murine leukemia virus DNA *in vitro*. *Journal of Virology* **64**, 5645-5648.

Bushman, F.D., Craigie, R., (1991). Activites of human immunodeficiency virus (HIV) integration protein in vitro: Specific cleavage and integration of HIV DNA. *Proceedings of the National Academy of Sciences USA* **88**, 1339-1343.

Bushman, F.D., Craigie, R., (1992). Integration of human immunodeficiency virus DNA: Adduct interference analysis of required DNA sites. *Proceedings of the National Academy of Sciences USA* **89**, 3458-3462.

Bushman, F.D., Engelman, A., Palmer, I., Wingfield, P., Craigie, R., (1993). Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynuclotidyl transfer and zinc binding. *Proceedings of the National Academy of Sciences USA* **90**, 3428-3432.

Bushman, F.D., Fujiwara, T., Craigie, R., (1990). Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* **249**, 1555-1558.

Cármenes, R.S., Freije, J.P., Molina M.M., Martín, J.M., (1989). Predict7, a program for protein structure prediction. *Biochemical and Biophysical Research Communications* **159**, 687-693.

Carteau, S., Mouscadet, J.M., Goulaouic, H., Subra, F., Auclair, C., (1993). Quantitative *in vitro* assay for human immunodeficiency virus deoxyribonucleic acid integration. *Archives of Biochemistry and Biophysics* **300**, 756-760.

Chatis, P.A., Holland, C.A., Hartley, J.W., Rowe, W.P., Hopkins, N., (1983). role for the 3' end of the genome in determining specificity of Friend and Moloney murine leukemia viruses. *Proceedings of the National Academy of Sciences USA* **80**, 4408-4411.

Cheng, Y-S.E., (1983). Increased cell buoyant densities of protein overproducing *Escherichia coli* cells. *Biochemical and Biophysical Research Communications* **111**, 104-111.

Chow, S.A., Vincent, K.A., Ellison, V., Brown, P.O., (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **255**, 723-726.

Cobrinik, D., Aiyar, A., Ge, Z., Katzman, M., Huang, H., Leis, J., (1991). Overlapping retrovirus U5 sequence elements are required for effecient integration and initiation of reverse transcription. *Journal of Virology* **65**, 3864-3872.

Cobrinik, D., Katz, R., Terry, R., Skalka, A.M., Leis, J., (1987). Avian sarcoma and leukosis virus *pol*-endonuclease recognition of the tandem long terminal repeat junction: Minimum site required for cleavage is also required for viral growth. *Journal of Virology* **61**, 1999-2008.

Coffin, J.M., (1990). Retroviridae and their replication. p. 1437-1500. *In* B.N. Fields and D.M. Knipe (ed.), Virology. Raven Press, New York.

Coleman, J.E., (1992). Zinc proteins: Enzymes, storage proteins, transcription factors, and replication proteins. *Annual Reviews in Biochemistry* **61**, 897-946.

Colicelli, J., Goff, S.P., (1985). Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* **42**, 573-580.

Colicelli, J., Goff, S.P., (1988). Isolation of an integrated provirus of Moloney murine leukemia virus with long terminal repeats in inverted orientation: Integration utilizing two U3 sequences. *Journal of Virology* **62**, 633-636.

Colicelli, J., Goff, S.P., (1988a). Sequence and spacing requirements of a retrovirus integration site. *Journal of Molecular Biology* **199**, 47-59.

Craigie, R., (1992). Hotspots and warm spots: Integration specificity of retroelements. *Trends in Genetics* **8**, 187-189.

Craigie, R., Fujiwara, T., Bushman, F., (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* **62**, 829-837.

Craigie, R., Mizuuchi, K., Bushman, F.D., Engelman, A., (1991). A rapid in vitro assay for HIV DNA integration. *Nucleic Acids Research* **19**, 2729-2734.

Cuypers, H.T., Selten, G., Quint, W., Zijlstra, M., Maandag, E.R., Boeleus, W., van Wezenbeck, P., Melief, C., Berns, A., (1984). Murine leukemia virus-induced T-cell lymphomagenesis: Integration of proviruses in a distinct chromosomal region. *Cell* **37**, 141-150.

Daar, E.S., Ho, D.D., (1990). The structure and function of retroviral envelope glycoproteins. *Seminars in Virology* **1**, 205-214.

Davis, B.R., Brightman, B.K., Chandy, K.G., Fan, H., (1987). Characterization of a preleukemic state induced by Moloney murine leukemia virus: Evidence for two infection events during leukomogenesis. *Proceedings of the National Academy of Sciences USA* **84**, 4875-4879.

Deen, C., Claassen, E., Gerritse, K., Zegers, N.D., Boersma, W.J.A., (1990). A novel carbodiimide coupling method for synthetic peptides. *Journal of Immunological Methods* **129**, 119-125.

Dhar, R., McClements, W.L., Enquist, L.W., Vande Woude G.F., (1980). Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proceedings of the National Academy of Sciences USA* **77**, 3937-3941.

Donehower, L.A., (1988). Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3' region of *pol*. *Journal of Virology* **62**, 3958-3964.

Doolittle, R.F., Feng, D.F., Johnson, M.S., McClure, M.A., (1989). Origins and evolutionary relationships of retroviruses. *Quarterly Review of Biology* **64**, 1-30.

Drelich, M., Wilhelm, R., Mous, J., (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. *Virology* **188**, 459-468.

Ellis, J., Bernstein, A., (1989). Retrovirus vectors containing an internal attachement site: Evidence that circles are not intermediates to murine retrovirus integration. *Journal of Virology* **63**, 2844-2846.

Ellison, V., Abrams, H., Roe, T., Lifson, J., Brown, P., (1990). Human immunodeficiency virus integration in a cell-free system. *Journal of Virology* **64**, 2711-2715.

Emini, E.A., Hughes J.V., Perlow, D.S., Boger, J., (1985). Induction of hepatitis A virus-neutralizing antibody by a virus specific synthetic peptide. *Journal of Virology* **55**, 836-839.

Engelman, A., Bushman, F.D., Craigie, R., (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO* **12**, 3269-3275.

Engelman, A., Craigie, R., (1992). Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *Journal of Virology* **66**, 6361-6369.

Engelman, A., Mizuuchi, K., Criagie, R., (1991). HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**, 1211-1221.

Evans, R.M., Hollenberg, S.M., (1988). Zinc fingers: Gilt by association. *Cell* **52**, 1-3.

Fan, H., (1990). Influences of the long terminal repeats on retrovirus pathogenicity. *Seminars in Virology* **1**, 165-174.

Farnet, C.M., Haseltine, W.A., (1991a). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *Journal of Virology* **65**, 1910-1915.

Farnet, C.M., Haseltine, W.A., (1991b). Circularization of human immunodeficiency virus type 1 DNA in vitro. *Journal of Virology* **65**, 6942-6952.

Felsenstein, K.M., Goff, S.P., (1992). Mutational analysis of the *gag-pol* junction of Molongy murine leukemia virus: Requirements for expression of the *gag-pol* fusion protein. *Journal of Virology* **66**, 6601-6608.

Feng, Y-X., Yuan, H., Rein, A., Levin, J.G., (1992). Bipartite signal for read-through suppression in murine leukemia virus mRNA: An eight-nucleotide purine-rivh sequence immediately downstream of the *gag* termination codon followed by an RNA pseudoknot. *Journal of Virology* **66**, 5127-5132.

Feng, Y.X., Copeland, T.D., Oroszlan, S., Rein, A., Levin, J.G., (1990). Identification of amino acids inserted during suppression of UAA and UGA termination codons at the *gag-pol* junction of Moloney murine leukemia virus. *Proceedings of the National Academy of Sciences USA* **87**, 8860-8863.

Fischer, B., Sumner, I., Goodenough, P., (1993) Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion bodies. *Biotechnology and Bioengineering* **41**, 3-13.

Fitzgerald, M.L., Vora, A.C., Grandgenett, D.P., (1991). Development of an acid-soluble assay for measuring retrovirus integrase 3'-OH terminal nuclease activity. *Analytical Biochemistry* **196**, 19-23.

Fitzgerald, M.L., Vora, A.C., Zeh, W.G., Grandgenett, D.P., (1992). Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. *Journal of Virology* **66**, 6257-6263.

Fujiwara, T., Craigie, R., (1989). Integration of mini-retroviral DNA: A cell-free reaction for biochemical analysis of retroviral integration. *Proceedings of the National Academy of Sciences USA* **86**, 3065-3069.

Fujiwara, T., Mizuuchi, K., (1988). Retroviral DNA integration: Structure of an integration intermediate. *Cell* **54**, 497-504.

Garfinkel, D.J., Hedge, A-M., Youngren, S.D., Copeland, T.D., (1991). Proteolytic processing of *pol-TYB* proteins from the yeast retrotransposon Ty1. *Journal of Virology* **65**, 4573-4581.

Garnier, J., Osguthorpe, D.J., Robson, B., (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* **120**, 97-120.

Gething, M-J., Sambrook, J., (1992). Protein folding in the cell. *Nature* **355**, 33-45.

Geysen, H.M., Tainer, J.A., Rodda, S.J., Mason, T.J., Alexander, H., Getzoff, E.D., Lerner, R.A., (1987). Chemistry of antibody binding to a protein. *Science* **235**, 1184-1190.

Goff, S.A., Goldberg, A.L., (1985). Production of abnormal proteins in *E. coli* stimulates transcription of *lon* and other heat shock genes. *Cell* **41**, 587-595.

Goff, S.P., (1990). Integration of retroviral DNA into the genome of the infected cell. *Cancer Cells* **2**, 172-178.

Goff, S.P., (1992). Genetics of retroviral integration. *Annual Reviews in Genetics* **26**, 527-544.

Golemis, E.A., Speck, N.A., Hopkins, N., (1990). Alignment of U3 region sequences of mammalian Type C viruses: Identification of highly conserved motifs and implications for enhancer design. *Journal of Virology* **64**, 534-542.

Grandgenett, D.P., Inman, R.B., Vora, A.C., Fitzgerald, M.L., (1993). Comparison of DNA binding and integration half-site selection by avian myeloblastosis virus integrase. *Journal of Virology* **67**, 2628-2636.

Grandgenett, D.P., Vora, A.C., (1985). Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. *Nucleic Acids Research* **13**, 6205-6221.

Grandgenett, D.P., Vora, A.C., Swanstrom, R., Olsen, J.C., (1986). Nuclease mechanism of the avian retrovirus pp32 endonuclease. *Journal of Virology* **58**, 970-974.

Gribskov, M., Burgess, R.R., (1983). Overexpression and purification of the sigma subunit of *Escherichia coli* RNA polymerase. *Gene* **26**, 109-118.

Harley, D.L., Kane, J.F., (1988). Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochemical Society Transactions* **16**, 101-102.

Hizi, A., Hughes, S.H., (1988). Expression of the Moloney murine leukemia virus and human immunodeficiency virus integration proteins in *Escherichia coli*. *Virology* **167**, 634-638.

Hong, T., Murphy, E., Groarke, J., Drlica, K., (1993). Human immunodeficiency virus type-1 DNA integration: Fine structure analysis using synthetic oligonucleotides. *Journal of Virology* **67**, 1127-1131.

Hopp, T.P., Woods, K.R., (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences USA* **78**, 3824-3828.

Hopp, T.P., Woods, K.R., (1983). A computer program for predicting protein antigenic determinants. *Molecular Immunology* **20**, 483-489.

Horton, R., Mumm, S., Grandgenett, D.P., (1988). Avian retrovirus pp32 DNA endonuclease is phosphorylated on Ser in the carboxyl-terminal region. *Journal of Virology* **62**, 2067-2075.

Horton, R., Mumm, S.R., Grandgenett, D.P., (1991). Phosphorylation of the avian retrovirus integration protein and proteolytic processing of it's carboxyl terminus. *Journal of Virology* **65**, 1141-1148.

Housset, V., DeRocquigny, H., Roques, B.P., Darlix, J.L., (1993). Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *Journal of Virology* **67**, 2537-2545.

Hu, S.C., Court, D.L., Zweig, M., Levin, J.G., (1986). Murine leukemia virus *pol* gene products: Analysis with antisera generated against reverse transcriptase and endonuclease fusion proteins expressed in *Escherichia coli*. *Journal of Virology* **60**, 267-274.

Hu, W-S., Temin, H.M., (1990). Retroviral recombination and reverse transcription. *Science* **250**, 1227-1233.

Hughes, S.H., Vogt, P.K., Stubblefield, E., Bishop, J.M., Varmus, H.E., (1981). Integration of avian sarcoma virus DNA in chicken cells. *Virology* **108**, 208-221.

Hunter, E., Swanstrom, R., (1990). Retrovirus envelope glycoproteins. p. 187-253. *In* R. Swanstrom and P.K. Vogt (ed.), Current Topics in Microbiology and Immunology: Retroviruses, Strategies of Replication, Vol. 157. Springer-Verlag, Berlin.

Ishimoto, L.K., Halperin, M., Champoux, J.J., (1991). Moloney murine leukemia virus IN protein from disrupted virions binds and specifically cleaves its target sequence *in vitro*. *Virology* **180**, 527-534.

Jaenicke, R., (1991). Protein folding: Local structures, domains, subunits, and assemblies. *Biochemistry* **30**, 3147-3161.

Jaenicke, R., (1993). Role of accessory proteins in protein folding. *Current Opinion in Structural Biology* **3**, 104-112.

Janin, J., Wodak, S., (1978). Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology* **125**, 357-386.

Johnson, M.S., McClure, M.A., Feng, D-F., Gray, J., Doolittle, R.F., (1986). Computer analysis of retroviral *pol* genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proceedings of the National Academy of Sciences USA* **83**, 7648-7652.

Johnston, B.A., Eisen, J., Fry, D., (1991). An evaluation of several adjuvant emulsion regimens for the production of polyclonal antisera in rabbits. *Laboratory Animal Science* **41**, 15-21.

Jones, K.S., Coleman, J., Merkel, G.W., Laue, T.M., Sklaka, A.M., (1992). Retroviral integrase functions as a multimer and can turn over catalytically. *Journal of Biological Chemistry* **267**, 16037-16040.

Jonsson, C.B., Donzella, G.A., Roth, M.J., (1993). Characterization of the forward and reverse integration reactions of the Moloney murine leukemia virus integrase protein purified from *Escherichia coli*. *Journal of Biological Chemistry* **268**, 1462-1469.

Jonsson, C.B., Roth, M.J., (1993). Role of the His-Cys finger of Moloney murine leukemia virus integrase protein in integration and disintegration. *Journal of Virology* **67**, 5562-5571.

Karplus, P.A., Schulz, G.E., (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften* **72**, 212-213.

Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T., Oroszlan, S., (1985). Murine leukemia virus maturation: Protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145**, 280-292.

Katz, R.A., Mack, J.P.G., Merkel, G., Kulkosky, J., Ge, Z., Leis, J., Skalka, A.M., (1992). Requirement for a conserved serine in both processing and joining activities of retroviral integrase. *Proceedings of the National Academy of Sciences USA* **89**, 6741-6745.

Katz, R.A., Merkel, G., Kulkosky, J., Leis, J., Skalka, A.M., (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell* **63**, 87-95.

Katzman, M., Katz, R.A., Skalka, A.M., Leis, J., (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the *in vivo* sites of integration. *Journal of Virology* **63**, 5319-5327.

Katzman, M., Mack, J.P.G., Skalka, A.M., Leis, J., (1991). A covalent complex between retroviral integrase and nicked substrate DNA. *Proceedings of the National Academy of Sciences USA* **88**, 4695-4699.

Keohavong, P., Thilly, W.J., (1989). Fidelity of DNA polymerases in DNA amplification. *Proceedings of the National Academy of Sciences* **86**, 9253-9257.

Khan, E., Mack, J.P.G., Katz, R.A., Kulkosky, J., Skalka, A.M., (1991). Retroviral integrase domains: DNA binding and

the recognition of LTR sequences. *Nucleic Acids Research* **19**, 851-860.

Kim, J.W., Closs, E.I., Albritton, L.M., Cunningham, J.M., (1991). Transport of cationic amino acids by the mouse ecotropic retrovirus receptor. *Nature* **352**, 725-728.

Kinsman, R.G., Olivier, G.W.J. Counter-ion distribution monitoring: A critical appraisal. In press.

Kitamura, Y., Ha Lee, Y.M., Coffin, J.M., (1992). Nonrandom integration of retroviral DNA *in vitro*: Effect of CpG methylation. *Proceedings of the National Academy of Sciences USA* **89**, 5532-5536.

Knappik, A., Krebber, C., Plückthun, A., (1993). The effect of folding catalysts on the *in vivo* folding process of different antibody fragments expressed in *Escherichia coli*. *Bio/Technology* **11**, 77-83.

Krogstad, P.A., Champoux, J.J., (1990). Sequence-specific binding of DNA by the Moloney murine leukemia virus integrase protein. *Journal of Virology* **64**, 2796-2801.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G., Skalka, A.M., (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Molecular and Cellular Biology* **12**, 2331-2338.

Kulkosky, J., Katz, R.A., Skalka, A.M., (1990). Terminal nucleotides of the preintegrative linear form of HIV-1 DNA deduced from the seqeunce of circular DNA junctions. *Journal of Acquired Immune Deficiency Syndrome* **3**, 852-858.

Kulkosky, J., Skalka, A.M., (1990). HIV DNA integration: Observations and inferences. *Journal of Acquired Immune Deficiency Syndrome* **3**, 839-851.

Kyte, J., Doolittle, R.F., (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105-132.

LaFemina, R.L., Callahan, P.L., Cordingly, M.G., (1991). Substrate specificity of recombinant human immunodeficiency virus integrase protein. *Journal of Virology* **65**, 5624-5630.

LaFemina, R.L., Schneider, C.L., Robbins, H.L., Callahan, P.L., LeGrow, K., Roth, E., Schleif, W.A., Emini, E.A., (1992). Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. *Journal of Virology* **66**, 7414-7419.

Leavitt, A.D., Shine, L., Varmus, H.E., (1993). Site-directed mutagenesis of HIV-1 integrase demonstrates

differential effects on integrase functions in vitro. *Journal of Biological Chemistry* **268**, 2113-2119.

Lerner, R.A., Green, N., Alexander, H., Liu, F-T., Sutcliffe, J.G., Shinnick, T.M., (1981). Chemically synthesized peptidespredicted from the nucleotide sequence of the hepatitis B virus genome elicit antibodies reactive with the native envelope protein of Dane particles. *Proceedings of the National Academy of Sciences USA* **78**, 3403-3407.

Leskowitz, S., Waksman B.H., (1960). Studies in immunization. I. The effect of route of injection of bovine serum albumin in Freund adjuvant on production of circulating antibody and delayed hypersensitivity. *Journal of Immunology* **84**, 58-72.

Li, G., Simm, M., Potash, M.J., Bolsky, D.J., (1993). Human immunodeficiency virus type 1 DNA synthesis, integration, and efficient viral replication in growth-arrested T-cells. *Journal of Virology* **67**, 3969-3977.

Lin, L-N., Hasumi, H., Brandt, J.F., (1988). Catalysis of proline isomerization during protein-folding reactions. *Biochimica et Biophysica Acta* **956**, 256-266.

Lin, T-H., Grandgenett, D.P., (1991). Retrovirus integrase: Identification of a potential leucine zipper motif. *Protein Engineering* **4**, 435-441.

Lineal, M.L., Miller, A.D., (1990). Retroviral RNA packaging: Sequence requirements and implications. p. 125-152. *In* R. Swanstrom and P.K. Vogt (ed.), Current Topics in Microbiology and Immunology: Retroviruses, Strategies of Replication. Vol. 157. Springer-Verlag, Berlin.

Lobel, L.I., Murphy, J.E., Goff, S.P., (1989). The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. *Journal of Virology* **63**, 2629-2637.

Lori, F., di Marzo Veronese, F., DeVico, A.L., Lusso, P., Reitz, M.S., Gallo R.C., (1992). Viral DNA carried by human immunodeficiency virus type 1 virions. *Journal of Virology* **66**, 5067-5074.

Luk, K-C., Gilmore T.D., Panganiban, A.T., (1987). The spleen necrosis virus *int* gene product expressed in *Escherichia coli* has DNA binding activity and mediates *att* and U5-specific DNA multimer formation *in vitro*. *Virology* **157**, 127-136.

Majors, J., (1990). The structure and function of retroviral long terminal repeats. p. 49-92. *In* R. Swanstrom and P.K. Vogt (ed.), Current Topics in Microbiology and Immunology: Retroviruses, Strategies of Replication, Vol. 157. Springer-Verlag, Berlin.

Makris, A., Patriotis, C., Bear, S.E., Tsichlis, P.N., (1993). Structure of a Moloney murine leukemia virus-like 30 recombinant: Implication for transduction of the c-Ha-*ras* proto-oncogene. *Journal of Virology* **67**, 1286-1291.

Maniatis, T., Fritsch, E.F., Sambrook, J., (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press. New York.

Marcus-Sekura, C.J., Woerner, A.M., Zweig, M., Court, D.L., Levin, J.G., Klutch, M., (1990). Expression of HIV-1 integrase in *E. coli*: Immunological analysis of the recombinant proteins. *AIDS Research and Human Retroviruses* **6**, 1399-1408.

Marczinovits, I., Molnár, J., Sóki, J., Fodor, I., (1992). Overexpression and purification of enzymatically active recombinant integrase protein of Rous sarcoma virus. *Virus Genes* **6**, 301-306.

Marston, F.A.O., (1986). The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. *Biochemical Journal* **240**, 1-12.

Miller, J., McLachlan, A.D., Klug, A., (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO* **4**, 1609-1614.

Mizuuchi, K., Craigie, R., (1986). Mechanism of bacteriophage Mu transposition. *Annual Reviews of Genetics* **20**, 385-429.

Moloney, J.B., (1960). Biological studies on a lymphoid-leukemia virus extracted from sarcoma 37. 1. Origin and introductory investigations. *Journal of the National Cancer Institute* **24**, 933-947.

Mooslehner, K., Karls, U., Harbers, K., (1990). Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *Journal of Virology* **64**, 3056-3058.

Muller, S., (1988). Peptide-carrier conjugation. p. 95-130. *In* M.H.V.van Regenmortel, J.P. Briand, S. Muller, S. Plaue (eds.), Laboratory Techniques in Biochemistry and Molecular Biology, vol 19. Synthetic Peptides as Antigens. Elsevier, Amsterdam.

Mumm, S.R., Grandgenett, D.P., (1991). Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis. *Journal of Virology* **65**, 1160-1167.

Mumm, S.R., Hippenmeyer, P.J., Grandgenett, D.P., (1992). Characterization of a stable eukaryotic cell line expressing the Rous sarcoma virus integrase. *Virology* **189**, 500-510.

Murphy, J.E., de Los Santos, T., Goff, S.P., (1993). Mutational analysis of the sequences at the termini of the Moloney murine leukemia virus DNA required for integration. *Virology* 195, 432-440.

Nash, M.A., Meyer, M.K., Decker, G.L., Arlinghaus R.B., (1993). A subset of $Pr65^{gag}$ is nucleus associated in murine leukemia virus-infected cells. *Journal of Virology* 67, 1350-1356.

Noguchi, T., (1991) Mini-maxi prep: an efficient and inexpensive procedure for medium scale preparation of high-quality supercoiled plasmid DNA. *Technique-A Journal of Methods in Cell and Molecular Biology* 3, 33-36.

Panet, A., Baltimore, D., (1987). Characterization of endonuclease activities in Moloney murine leukemia virus and its replication-defective mutants. *Journal of Virology* 61, 1756-1760.

Panganiban, A.T., (1990). Retroviral reverse transcription and DNA integration. *Seminars in Virology* 1, 187-194.

Panganiban, A.T., Temin, H.M., (1984). The retrovirus *pol* gene encodes a product required for DNA integration: Identification of a retrovirus *int* locus. *Proceedings of the National Academy of Sciences USA* 81, 7885-7889.

Párraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T., Klevit, R.E., (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* 241, 1489-1492.

Pauza, C.D., (1990). Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* 179, 886-889.

Pavletich, N.P., Pabo, C.O., (1991). Zinc finer-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809-817.

Peng, C., Chang, N.T., Chang, T.W., (1991). Identification and characterization of human immunodeficiency virus type-1 *gag-pol* fusion protein in transfected mammalian cells. *Journal of Virology* 65, 2751-2756.

Prakash, K., Ranganathan, P.N., Mettus, R., Reddy, P., Srinivasan, A., Plotkin, S., (1992). Generation of deletion mutants of simian immunodeficiency virus incapable of proviral integration. *Journal of Virology* 66, 167-171.

Prats, A.C., Sarih, L., Gabus, C., Litvak, S., Keith, G., Darlix, J.L., (1988). Small finger protein of avain and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA. *EMBO* 7, 1777-1783.

Pryciak, P.M., Müller, H-P., Varmus, H.E., (1992). Simian virus 40 minichromosomes as targets for retroviral integration *in vivo*. *Proceedings of the National Academy of Sciences USA* **89**, 9237-9241.

Pryciak, P.M., Sil, A., Varmus, H.E., (1992). Retroviral integration into minichromosomes *in vitro*. *EMBO* **11**, 291-303.

Pullen, K.A., Champoux, J.J., (1990). Plus-strand origin for human immunodeficiency virus type 1: Implications for integration. *Journal of Virology* **64**, 6274-6277.

Rein, A., McClure, M.R., Rice, N.R., Luftig, R.B., Schultz, A.M., (1986). Myristylation site in $Pr65^{gag}$ is essential for virus particle formation by Moloney murine leukemia virus. *Proceedings of the National Academy of Sciences USA* **83**, 7246-7250.

Rhee, S.S., Hunter, E., (1990). Structural role of the matrix protein of type D retroviruses in Gag polyprotein stability and capsid assembly. *Journal of Virology* **64**, 4383-4389.

Riniker, B., Hartman, A., (1990). Deprotection of peptides containing Arg(Pmc) and tryptophan or tyrosine: Elucidation of byproducts. *Peptides, Chemistry, Structure and Biology: Proceedings of the 11th APS,* 950-952.

Roe, T.Y., Reynolds, T.C., Yu, G., Brown, P.O., (1993). Integration of murine leukemia virus DNA depends on mitosis. *EMBO* **12**, 2099-2108.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., Breindl, M., (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNaseI-hypersensitive sites. *Journal of Virology* **61**, 336-343.

Roth, M.J., (1991). Mutational analysis of the carboxyl terminus of the Moloney murine leukemia virus integration protein. *Journal of Virology* **65**, 2141-2145.

Roth, M.J., Schwartzberg, P., Tanese, N., Goff, S.P., (1990). Analysis of mutations in the integration function of Moloney murine leukemia virus: Effects on DNA binding and cutting. *Journal of Virology* **64**, 4709-4717.

Roth, M.J., Schwartzberg, P.L., Goff, S.P., (1989). Structure of the termini of DNA intermediates in the integration of retroviral DNA: Dependence on IN function and terminal DNA sequence. *Cell* **58**, 47-54.

Roth, M.J., Tanese, N., Goff, S.P., (1988). Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. *Journal of Molecular Biology* **203**, 131-139.

Sakai, H., Kawamura, M., Sakuragi, J-I., Sakuragi, S., Shibata, R., Ishimoto, A., Ono, N., Ueda, S., Adachi, A., (1993). Integration is essential for efficient gene expression of human immunodeficiency virus type 1. *Journal of Virology* **67**, 1169-1174.

Salisbury, S.A., Tremeer, E.J., Davies, J.W., Owen, D.E.I.A., (1990). Acylation monitoring in solid phase peptide synthesis by the equilibrium distribution of coloured ions. *Journal of the Chemical Society Chemical Communications* **7**, 538-543.

Sanger, F., Nicklen, S., Coulson, R., (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**, 5463-5467.

Schauer, M., Billich, A., (1992). The N-terminal region of HIV-1 integrase is required for integration activity, but no for DNA-binding. *Biochemical and Biophysical Research Communications* **185**, 874-880.

Schein, C.H., (1989). Production of soluble recombinant proteins in bacteria. *Bio/Technology* **7**,, 1141-1149.

Schein, C.H., Noteborn, M.J.M., (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth termperature. *Bio/Technology* **6**, 291-294.

Scherdin, U., Rhodes, K., Breindl, M., (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *Journal of Virology* **64**, 907-912.

Schmid, F.X., (1993). Prolyl isomerase: Enzymatic catalysis of slow protein-folding reactions. *Annual Review of Biophysics and Biomolecular Structure* **22**, 123-143.

Schoemaker, J.M., Brasnett, A.H., Marston, F.A.O., (1985). Examination of calf prochymosin accumulation in *Escherichia coli*: Disulphide linkages are a structural component of prochymosin containing inclusion bodies. *EMBO* **4**, 775-780.

Sherman, P.A., Dickson, M.L., Fyfe, J.A., (1992). Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. *Journal of Virology* **66**, 3593-3601.

Sherman, P.A., Fyfe, J.A., (1990). Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selecive DNA cleaving activity. *Proceedings of the National Academy of Sciences USA* **87**, 5119-5123.

Shinnick, T.M., Lerner, R.A., Sutcliffe, J.G., (1981). Nucleotide sequence of Moloney murine leukaemia virus. *Nature* **293**, 543-548.

Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S.W., Baltimore, D., (1980). Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: Implications for retrovirus integration. *Proceedings of the National Academy of Sciences USA* **77**, 3932-3936.

Sieber, P., (1987). Modification of tryptophan residues during acidolysis of 4-methoxy-2,3,6-trimethylbenzene-sulfonyl groups, effects of sscavengers. *Tetrahedron Letters* **28**, 1637-1640.

Skalka, A.M., (1989). Retroviral proteases: First glimpse of the anatomy of a processing machine. *Cell* **56**, 911-913.

Smith, D.E., O'Brien, M.E., Palmer, V.J., Sadowski, J.A., (1992). The selection of an adjuvant emulsion for polyclonal antibody production using a low molecular weight antigen in rabbits. *Laboratory Animal Science* **42**, 599-601.

Speck, N.A., Baltimore, D., (1987). Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Molecular and Cellular Biology* **7**, 1101-1110.

Steffen, D., (1984). Proviruses are adjacent to c-*myc* in some murine leukemia virus-induced lymphomas. *Proceedings of the National Academy of Sciences USA* **81**, 2097-2101.

Stevenson, M., Stanwick, T.L., Dempsey, M.P. Lamonica, C.A., (1990). HIV-1 replication is controlled at the level of T-cell activation and proviral integration. *EMBO* **9**, 1551-1560.

Storch, T.G., Arnstein, P., Manohar, V., Leiserson, W.M., Chused, T.M., (1985). Proliferation of infected lymphoid precursers before Moloney murine leukemia virus-induced T-cell lymphoma. *Journal of the National Cancer Institute* **74**, 137-143.

Sutcliffe, J.G., Shinnick, T.M., Green, N., Lerner, R.A., (1983). Antibodies that react with predetermined sites on a protein. *Science* **219**, 660-666.

Tanese, N., Roth, M.J., Goff, S.P., (1986). Analysis of retroviral pol gene products with antisera raised against fusion proteins produced in *Esherichia coli*. *Journal of Virology* **59**, 328-340.

Temin, H.M., (1988). Mechanism of cell killing/cytopathic effects by nonhuman retroviruses. *Reviews of Infectious Diseases* **10**, 399-405.

Temin, H.M., Mizutani, S., (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211-1213.

Terry, R., Soltis, D.A., Katzman, M., Cobrinik, D., Leis, J., Skalka, A.M., (1988). Properties of avian sarcoma-leukosis virus pp32-related *pol*-endonucleases produced in *Escherichia coli*. *Journal of Virology* **62**, 2358-2365.

Tindall, K.R., Kunkel, T.A., (1988). Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27**, 6008-6013.

Trono, D., (1992). Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. *Journal of Virology* **66**, 4893-4900.

Tsichlis, P.N., (1987). Oncogenesis by Moloney murine leukemia virus. *Anticancer Research* **7**, 171-180.

van Beveren, C., Rands, E., Chattopadhyay, S.K., Lowy, D.R., Verma, I.M., (1982). Long terminal repeat of murine retroviral DNAs: Sequence analysis, host-proviral junction, and preintegration site. *Journal of Virology* **41**, 542-556.

van Gent, D.C., Elgersma, Y., Bolk, M.W.J., Vink, C., Plasterk, R.H.A., (1991). DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. *Nucleic Acids Research* **19**, 3821-3827.

van Gent, D.C., Oude Groeneger, A.A.M., Plasterk, R.H.A., (1992). Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proceedings of the National Academy of Sciences USA* **89**, 9598-9602.

van Gent, D.C., Oude Groeneger, A.A.M., Plasterk, R.H.A., (1993a). Identification of amino acids in HIV-1 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. *Nucleic Acids Research* **21**, 3373-3377.

van Gent, D.C., Vink, C., Oude Groeneger, A.A.M., Plasterk, R.H.A., (1993b). Complementation between HIV integrase proteins mutated in different domains. *EMBO* **12**, 3261-3267.

van Regenmortel, M.H.V., Briand, J.P., Muller, S., Plaue, S., (1988). Synthetic Polypeptides as Antigens. Vol. 19, Laboratory Techniques in Biochemistry and Molecular Biology. Elsevier, Amsterdam.

Varmus, H., Brown R., (1989). Retroviruses. p. 53-109. *In* D.E. Berg and M.M. Howe (ed.), Mobile DNA. American Society for Microbiology, Wash. D.C.

Varmus, H.E., (1982). Form and function of retroviral proviruses. *Science* **216**, 812-820.

Vijaya, S., Steffen, D.L., Robinson, H.L., (1986). Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. *Journal of Virology* **60**, 683-692.

Vincent, K.A., Ellison, V., Chow, S.A., Brown, P.O., (1993). Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. *Journal of Virology* **67**, 425-437.

Vincent, K.A., York-Higgens, D., Quiroga, M., Brown, P.O., (1990). Host sequences flanking the HIV provirus. *Nucleic Acids Research* **18**, 6045-6047.

Vink, C., Groenink, M., Elgersma, Y., Fouchier, R.A.M., Tersmette, M., Plasterk, R.H.A., (1990). Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA. *Journal of Virology* **64**, 5626-5627.

Vink, C., Oude Groeneger, A.A.M., Plasterk, R.H.A., (1993). Identificaiton of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein. *Nucleic Acids Research* **21**, 1419-1425.

Vink, C., van Gent, D.C., Elgersma, Y., Plasterk, R.H.A., (1991b). Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNa recognition sequence for efficient cleavage. *Journal of Virology* **65**, 4636-4644.

Vink, C., van Gent, D.C., Plasterk, R.H.R., (1990). Integration of human immunodeficiency virus types 1 and 2 DNA *in vitro* by cytoplasmic extracts of Moloney murine leukemia virus-infected mouse NIH 3T3 cells. *Journal of Virology*, **64**, 5219-5222.

Vink, C., Yeheskiely, E., van der Marel, G., van Boom, J.H., Plasterk, R.H.A., (1991a). Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. *Nucleic Acids Research* **19**, 6691-6698.

Vora, A.C., Fitzgerald, M.L., Grandgenett, D.P., (1990). Removal of 3'-OH-terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein. *Journal of Virology* **64**, 5656-5659.

Walter, G., (1986). Production and use of antibodies against synthetic peptides. *Journal of Immunological Methods* **88**, 149-161.

Wang, J., Kavenaugh, M.P., North, R.A., Kabat, D., (1991). Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. *Nature* **352**, 729-731.

Warren, H.S., Vogel, F.R., Chedid, L.A., (1986). Current status of immunological adjuvants. *Annual Reviews in Immunology* **4**, 369-388.

Welling, G.W., Weijer, W.J., van der Zee, R., Welling-Wester, S., (1985). Prediction of sequential antigenic regions in proteins. *FEBS Letters* **188**, 215-218.

Westall, F.C., Robinson, A.B., (1970). Solvent modification in Merrifield solid-phase peptide synthesis. *Journal of Organic Chemistry* **35**, 2842-2844.

Whitcomb, J.M., Hughes, S.H., (1992). Retroviral reverse transcription and integration: Progress and problems. *Annual Review in Cell Biology* **8**, 275-306.

White, P., (1992). Fmoc-Trp(Boc)-OH: A new derivative for the synthesis of peptides containing tryptophan. *Peptides, Chemistry, Structure and Biology: Proceedings of the 12th APS.* p. 537-538.

Wills, N.M., Gesteland, R.F., Atkins, J.F., (1991). Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus *gag* stop codon. *Proceedings of the National Academy of Sciences USA* **88**, 6991-6995.

Woerner, A.M., Marcus-Sekura, C.J., (1993). Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Research* **21**, 3507-3511.

Yoshinaka, Y., Katoh, I., Copeland, T.D., Oroszlan, S., (1985). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proceedings of the National Academy of Sciences USA* **82**, 1618-1622.

Yoshinaka, Y., Luftif, R.B., (1977). Murine leukemia virus morphogenesis: Cleavage of p70 *in vitro* can be accompanied by a shift from a concentrically coiled internal strand ("immature") to a collapsed ("mature") form of the virus core. *Proceedings of the National Academy of Science USA* **74**, 3446-3450.

Yoshinaka, Y., Luftig, R.B., (1978). Morphological conversion of 'immature' Rauscher leukaemia virus cores to a 'mature' form after addition of the p65-70 (*gag* gene product) proteolytic factor. *Journal of General Virology* **40**, 151-160.

Yourno, J., Kohno, T., Roth, J.R., (1970). Enzyme evolution: Generation of a bifunctional enzyme by fusion of adjacent genes. *Nature* **228**, 820-824.

Zervos, P., Hassell, T., Van Frank, R., Lai, M.T., (1991). Characterization of an internally initiated integrase protein of HIV-1 produced in *E. coli*. *Biochemical and Biophysical Research Communications* **170**, 1061-1066.

Vol. 189, No. 2, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
December 15, 1992 Pages 813-818

ZINC BINDING BY RETROVIRAL INTEGRASE

A.R. McEuen, B. Edwards, K.A. Koepke, A.E. Ball, B.A. Jennings, A.J. Wolstenholme, M.J. Danson, and D.W. Hough

Biochemistry Dept., University of Bath, Claverton Down, Bath BA2 7AY, U.K.

Received October 20, 1992

Zinc binding by integrase from Moloney murine leukaemia virus and a protein A fusion protein containing integrase from human immunodeficiency virus type 1 was demonstrated by a zinc blotting technique using $^{65}ZnCl_2$. Autoradiography revealed a clear band that was absent from the appropriate controls. This band co-migrated with the major band in Coomassie-stained gels and in immunoblots. This binding activity was retained in the presence of competing divalent cations and was sensitive to oxidation. This is the first demonstration of zinc binding by intact retroviral integrase.

A wide range of nucleic acid-binding proteins bind zinc as an essential cofactor (1-3). These employ a variety of structures ranging from the compact unit in retroviral nucleocapsid proteins (4), through the now classic zinc fingers of *Xenopus* TFIIIA, to the more complex structures found in GAL4 and glucocorticoid receptor (5). Integrase, the retroviral enzyme that integrates the dsDNA product of reverse transcription into the host genome (6, 7), was first identified as possessing a potential zinc-binding motif from an alignment of five sequences (8). Subsequent studies have shown this motif to be strongly conserved in integrases from 80 different retroviruses and retrotransposons (9). There are three notable differences between it and the canonical zinc finger of TFIIIA: 1) the sequence of ligands is HHCC rather than CCHH, 2) the loop between the second and third ligands is twice (*e.g.*, human immunodeficiency virus type 1 (HIV-1)) or nearly 3 times (*e.g.*, Moloney murine leukaemia virus (MoMLV)) as long, and 3) it is present only as a single copy, not a string of 3 - 12 similar motifs. Also, the pattern of conserved residues is different (Figure 1). Nonetheless, a role in DNA binding has been suggested (8).

Integrase acts on two different DNA molecules: the proviral DNA, from which it cleaves a dinucleotide at both 3' ends (10, 11), and the target or host DNA, into which it inserts the proviral DNA in a concerted reaction (12). The interaction with the first is sequence-specific, but that with the second is not. DNA-binding activity has been demonstrated by a variety of techniques (13-17). However, disruption of the putative zinc motif by deletion (9, 16, 17) or by

Abbreviations: MoMLV, Moloney murine leukaemia virus; HIV-1, human immunodeficiency virus type 1; SDS-PAGE, sodium dodecyl sulphate - polyacrylamide gel electrophoresis; DTPA, diethylenetriaminepentaacetic acid.

0006-291X/92 $4.00

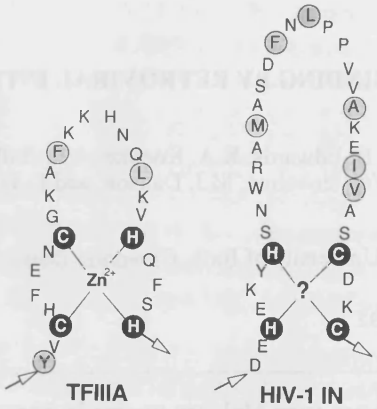


Figure 1. *Comparison of the retroviral motif with a classic zinc finger.* The structure on the left is the fourth zinc finger (res. 105-129) of Transcription Factor IIIA from *Xenopus laevis*. On the right is the putative zinc motif from HIV-1 integrase (res. 10-42). Black circles - Essential residues (100% conservation). Gray circles - highly conserved residues within the group under consideration; eukaryotic transcription factors(1, 3) or retroviral integrases (9).

substitution of one of the histidine (9) or cysteine (13) ligands did not noticeably affect DNA binding, although such mutations did reduce or abolish enzymic activity (9, 13, 16, 18). Rather, DNA-binding activity seemed to be localised to the C-terminal half of the protein (9, 15, 16).

In view of these results, the question arises as to whether the putative zinc motif of retroviral integrase actually does bind zinc. We sought to answer this by the method of zinc blotting. While this work was in progress, Burke *et al.* (19) published results showing that a 56 residue peptide from HIV-1 integrase containing this motif did indeed bind zinc. Our results here complement their findings by demonstrating zinc binding by the intact integrases from MoMLV and HIV-1.

MATERIALS AND METHODS

Expression and purification of proteins. As integrase is synthesized as a polyprotein, it lacks its own initiation codon. Therefore, the N-terminus must be modified for recombinant expression. The MoMLV integrase was expressed as a fusion with the three N-terminal residues of β-galactosidase (14). Plasmid pIF (gift from P. Krogstad) was transformed into *Escherichia coli* TG1. Cultures grown to mid-log phase in LB broth were induced with 1 mM isopropyl β-D-thiogalactopyranoside and harvested after a further 2.5 h growth at 32°C. Cells were resuspended in sonication buffer (50 mM Tris-HCl, 10 mM dithiothreitol, 1 mM diethylenetriaminepentaacetic acid (DTPA), pH 7.5), sonicated, and centrifuged. The pellet was extracted with 200 mM NaCl and 0.5% Nonidet P-40 in sonication buffer, and recentrifuged. Prior to electrophoresis, the pellet was taken up in sonication buffer containing 8 M urea, 0.1% sodium dodecyl sulphate (SDS) and 1.0% 2-mercaptoethanol.

The HIV-1 integrase was expressed as a fusion protein with staphylococcal protein A (20). A fragment of HIV-1 clone HB10 encoding the C-terminal 56 residues of reverse transcriptase and the entire integrase was excised from plasmid pOGS806 (gift from M. Richardson) using an engineered *Bam*H I site and a naturally occurring *Sal* I site, and ligated into pRIT5 (Pharmacia). The resultant plasmid, pINBE2, was cloned in *E. coli* CAG629 (developed by Dr. C. Cross, Univ. of Wisconsin). Cells were grown to stationary phase at 32°C and periplasmic proteins were removed by osmotic shock prior to sonication. The sonicated extract

was fractionated with ammonium sulphate, and the proteins which precipitated between 20% and 40% saturation at 4°C were resuspended in and dialysed against 50 mM Tris-HCl, 1 mM dithiothreitol, 0.1 mM DTPA, pH 7.5. After clarification by centrifugation, the supernatant was diluted with an equal volume of 50 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20, pH 7.6, and applied to a column of IgG-Sepharose (Pharmacia) following manufacturer's instructions (20). Bound proteins were eluted with 0.5 M ammonium acetate, pH 3.4. Fractions containing protein by A_{280} were neutralised with 1 M Tris base, pooled, and dialysed against 1 mM Tris-HCl, 20 μM dithiothreitol, 2 μM DTPA, pH 7.5. Samples were concentrated by freeze-drying and reconstituted in sonication buffer.

Zinc binding. The zinc blotting procedure of Barbosa *et al.* (21), which includes 10 mM dithiothreitol in the renaturation buffer, was employed. Nitrocellulose blots were incubated in $^{65}ZnCl_2$ (2.4 Ci/g, 50 μM) in N_2-purged labelling buffer for 30 min, rinsed with three changes of buffer containing 1 mM dithiothreitol, and the zinc-binding proteins visualised by autoradiography.

Immunoblotting. After electrophoresis and transfer to nitrocellulose, MoMLV integrase was identified with rabbit serum raised against a synthetic peptide corresponding to residues 28 - 42 of the enzyme. Antibodies to *E. coli* proteins were removed by incubation overnight at 4°C with a nitrocellulose blot of the whole cell lysate of induced *E. coli* TG1 cells containing the parent expression vector pTZ19R. The HIV-1 fusion protein was identified with normal rabbit serum, which bound strongly with the protein A portion of the fusion protein and showed negligible reaction with host proteins at the dilution used (1:6000).

RESULTS AND DISCUSSION

Expression and purification of proteins. The MoMLV construct was expressed as inclusion bodies which remained insoluble even in the presence of 8 M urea. Only by the further addition of SDS and heating could they be solubilised. The HIV-1 construct was more soluble, up to 50% of the expressed fusion protein remained in solution, depending on growth conditions and host strain. Highest solubility was obtained by expression in *E. coli* CAG629 grown at 32°C. This fusion protein was retained in the cytoplasm even though it contained a signal peptide. Similar to what others have reported with the pRIT5 system (22, 23), there are a number of expressed products that are shorter than the full length fusion protein, but still possess the protein A moiety. The most abundant of these, equal in size to the protein A domain alone, was secreted to the periplasmic space, so an osmotic shock step was included to separate this fragment from the desired product. The presence of the integrase domain greatly reduced the affinity of the fusion protein for IgG, so that most of the fusion protein applied to the affinity column washed through without binding. Such an effect by the heterologous protein on the affinity tag has also been observed with fusions of portions of feline immunodeficiency virus *env* gene product with glutathione S transferase (G. Reid, personal communication). This problem was partially overcome by rechromatography.

Zinc blotting. When the MoMLV sample was assayed by this method, a strong band of radioactive zinc appeared at the same molecular mass (45 kDa) as the major protein band revealed by Coomassie staining and the major immunoreactive band detected with anti-peptide serum (Figure 2). No such bands were observed in the zinc blot, Coomassie-stained gel, or immunoblot of the insoluble fraction of host cells containing the parent vector pTZ19R. As normally soluble proteins may become trapped in insoluble inclusion bodies, whole cell lysate of

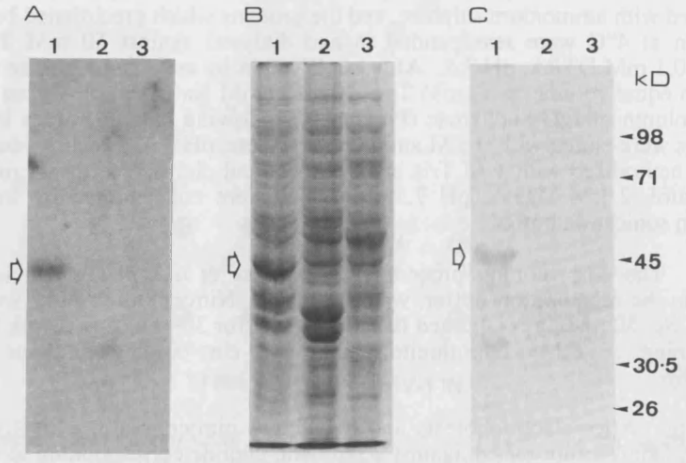


Figure 2. *Zinc binding by MoMLV integrase.* All samples were subjected to SDS-PAGE (8.0% gels). A) Zinc blot, B) Coomassie - stained gel, C) Immunoblot. In A, B, and C: 1) MoMLV integrase in insoluble fraction after sonication of pIF expressed in *E. coli* TG1, 2) insoluble fraction after sonication of parent vector pTZ19R expressed in *E. coli* TG1, 3) whole cell lysate of pTZ19R in *E. coli* TG1. Open arrow indicates MoMLV integrase. Positions of prestained marker proteins are shown on the right.

the host strain with pTZ19R was used as an additional control. No zinc-binding or immunoreactive bands were seen. Therefore, the observed zinc-binding activity can be attributed to the MoMLV integrase.

Similar results were obtained with the HIV-1 fusion protein (Figure 3), although, as noted above, a family of bands was observed rather than a single band in the zinc blot, Coomassie-

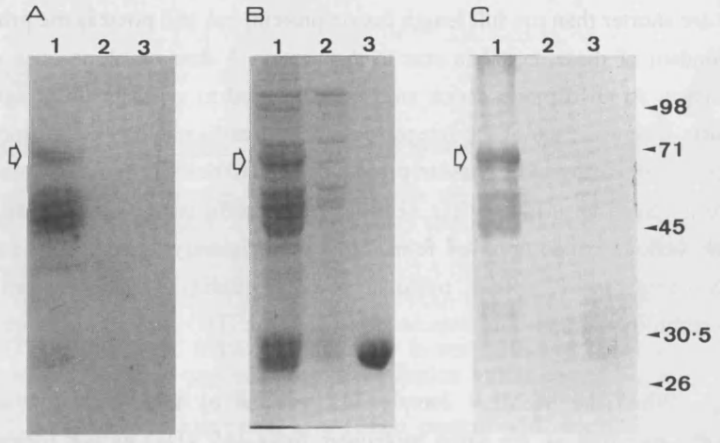


Figure 3. *Zinc binding by HIV-1 integrase.* All samples were subjected to SDS-PAGE (8.0% gels). A) Zinc blot, B) Coomassie - stained gel, C) Immunoblot. In A, B, and C: 1) HIV-1 integrase in a protein A-RT-IN fusion protein, 2) extract of host cells expressing parent plasmid subjected to same purification protocol as in lane 1, 3) purified staphylococcal protein A. Open arrow indicates protein A-RT-IN fusion protein. Positions of prestained marker proteins are shown on the right.

stained gel, and immunoblot. The slowest moving band has a molecular mass (71 kDa) expected of the full length fusion protein. This family of bands was absent from an extract of host cells expressing the parent plasmid pRIT5 subjected to the same purification protocol as the fusion protein. Also, the protein A moiety alone (purified by affinity chromatography from the osmotic shock fraction of *E. coli* cells expressing pRIT5) failed to bind any detectable zinc. The observed zinc-binding activity can therefore be confidently attributed to the integrase portion of the fusion protein. Although the evidence presented here does not preclude the possibility of zinc being bound by the C-terminal residues of reverse transcriptase, a complete metal binding site has not been observed in the crystal structure of this portion of the enzyme (24).

For both enzymes, zinc binding was lost if reducing conditions were not maintained during renaturation, a finding which supports the involvement of the conserved cysteines of the putative zinc motif. As integrase also requires Mn^{2+} for activity (Ca^{2+} and Mg^{2+} can substitute to a lesser degree)(9, 15, 16, 18), the possibility exists that $^{65}Zn^{2+}$ was binding adventitiously to the Mn^{2+} site. However, no diminution of ^{65}Zn-binding activity was seen in the presence of 10 mM Mn^{2+}, Ca^{2+}, or Mg^{2+} (a 200-fold molar excess over Zn^{2+}). In contrast, binding of ^{65}Zn was greatly reduced in the presence of 1 mM non-labelled $ZnCl_2$. Zinc, therefore, seems to be binding to a distinct site.

The results presented here, together with the strong conservation of this motif in retroviral integrases (9), and the proven ability of an isolated peptide containing this motif to bind zinc (19), strongly suggest that integrase binds a metal ion with this motif *in vivo*. With this added evidence that a zinc-containing structure does exist, the quest for its function can proceed with increased confidence. A role which does not seem to have been considered is a structural role independent of DNA binding as is seen in aspartate transcarbamylase (ATCase) (3). In this enzyme the integrity of the Zn-binding domain is essential for the correct association of the regulatory and catalytic subunits. The spacing of the zinc ligands in ATCase

$$\text{-Cys-}X_4\text{-Cys-}X_{25}\text{-Cys-}X_2\text{-Cys-}$$

is very similar to the spacing in HIV-1 integrase

$$\text{-His-}X_3\text{-His-}X_{23}\text{-Cys-}X_2\text{-Cys-}.$$

However, rather than forming a zinc finger, the intervening loop in ATCase contributes 2 strands to a 4-stranded β-structure (25). Similarly, the zinc motif of retroviral integrases could take up a tertiary structure totally unrelated to canonical zinc fingers.

ACKNOWLEDGMENTS

We would like to thank Dr. Paul Krogstad, Univ. of Washington, for the gift of plasmid pIF, Dr. Mark Richardson, British Biotechnology Ltd., for plasmid pOGS806, Dr. Sarah Cox, Univ. of Oxford, for *E. coli* CAG629, used by permission from Celltech, and Dr. Richard Kinsman, Univ. of Bath, for the synthetic peptides. This work was supported by a Wellcome Prize studentship to K.A.K. and by the MRC AIDS Directed Programme.

REFERENCES

1. Klug, A., and Rhodes, D. (1987) Trends Biochem. Sci. 12, 464 - 469.
2. Berg, J.M. (1990) Ann. Rev. Biophys. Biophys. Chem. 19, 405 - 421.
3. Coleman, J.E. (1992) Ann. Rev. Biochem. 61, 897 - 946.
4. Summers, M.F., Henderson, L.E., Chance, M.R., Bess, J.W., Jr., South, T.L., Blake, P.R., Sagi, I., Perez-Alvarado, G., Sowder, R.C., III, Hare, D.R., and Arthur, L.O. (1992) Protein Sci. 1, 563 - 574.
5. Luisi, B. (1992) Nature 356, 379 - 380.
6. Bushman, F.D., Fujiwara, T., and Craigie, R. (1990) Science 249, 1555 - 1558.
7. Craigie, R., Fujiwara, T., and Bushman, F. (1990) Cell 62, 829 - 837.
8. Johnson, M.S., McClure, M.A., Feng, D.-F., Gray, J., and Doolittle, R.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7648 - 7652.
9. Khan, E., Mack, J.P.G., Katz, R.A., Kulkosky, J., and Skalka, A.M. (1991) Nucl. Acid Res. 19, 851 - 860.
10. Fitzgerald, M.L., Vora, A.C., and Grandgenett, D.P. (1991) Anal. Biochem. 196, 19 - 23.
11. Vink, C., Yeheskiely, E., van der Marel, G.A., van Boom, J.H., and Plasterk, R.H.A. (1991) Nucl. Acid Res. 19, 6691 - 6698.
12. Engelman, A., Mizuuchi, K., and Craigie, R. (1991) Cell 67, 1211 - 1221.
13. Roth, M.J., Schwartzberg, P., Tanese, N., and Goff, S.P. (1990) J. Virol. 64, 4709 - 4717.
14. Krogstad, P.A., and Champoux, J.J. (1990) J. Virol. 64, 2796 - 2801.
15. Woerner, A.M., Klutch, M., Levin, J.G., and Marcus-Secura, C.J. (1992) AIDS Res. Human Retrovir. 8, 297 - 304.
16. Schauer, M., and Billich, A. (1992) Biochem. Biophys. Res. Comm. 185, 874 - 880.
17. Mumm, S.R., and Grandgenett, D.P. (1991) J. Virol. 65, 1160 - 1167.
18. Drelich, M., Wilhelm, R., and Mous, J. (1992) Virology 188, 459 - 468.
19. Burke, C.J., Sanyal, G., Bruner, M.W., Ryan, J.A., LaFemina, R.L., Robbins, H.L., Zeft, A.S., Middaugh, C.R., and Cordingley, M.G. (1992) J. Biol. Chem. 267, 9639 - 9644.
20. The Protein A System, Pharmacia, Molecular Biology Division, Uppsala.
21. Barbosa, M.S., Lowy, D.R., and Schiller, J.T. (1989) J. Virol. 63, 1404 - 1407.
22. Nilsson, B., Abrahmsen, L., and Uhlen, M. (1985) EMBO J. 4, 1075 - 1080.
23. Ducancel, F., Boulain, J.-C., Tremeau, O., and Menez, A. (1989) Protein Eng. 3, 139 - 143.
24. Davies, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R., and Matthews, D.A. (1991) Science 252, 88 - 95.
25. Honzatko, R.B., Crawford, J.L., Monaco, H.L., Ladner, J.E., Ewards, B.F.P., Evans, D.R., Warren, S.G., Wiley, D.C., Ladner, R.C. and Lipscomb, W.N. (1982) J. Mol. Biol. 160, 219 - 263.